

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of: **Mikio ITO et al.**

Serial Number: **Not Yet Assigned**

Filed: **December 19, 2003**

Customer No.: **38834**

For: **RAID APPARATUS AND LOGICAL DEVICE EXPANSION METHOD
THEREOF**

CLAIM FOR PRIORITY UNDER 35 U.S.C. 119

Commissioner for Patents
P. O. Box 1450
Alexandria, VA 22313-1450

December 19, 2003

Sir:

The benefit of the filing date of the following prior foreign application is hereby requested for the above-identified application, and the priority provided in 35 U.S.C. 119 is hereby claimed:

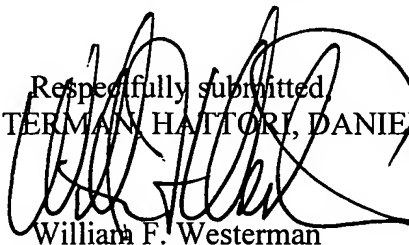
Japanese Appln. No. 2002-378284, filed on December 26, 2002

In support of this claim, the requisite certified copy of said original foreign application is filed herewith.

It is requested that the file of this application be marked to indicate that the applicants have complied with the requirements of 35 U.S.C. 119 and that the Patent and Trademark Office kindly acknowledge receipt of said certified copy.

In the event that any fees are due in connection with this paper, please charge our Deposit Account No. 50-2866.

Respectfully submitted,
WESTERMAN HATTORI, DANIELS & ADRIAN, LLP



William F. Westerman
Reg. No. 29,988

Atty. Docket No.: 032188
Suite 700
1250 Connecticut Avenue, N.W.
Washington, D.C. 20036
Tel: (202) 822-1100
Fax: (202) 822-1111
WFW/yap

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 2 年 1 2 月 2 6 日
Date of Application:

出 願 番 号 特 願 2 0 0 2 - 3 7 8 2 8 4
Application Number:
[ST. 10/C] : [J P 2 0 0 2 - 3 7 8 2 8 4]

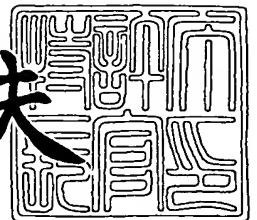
出 願 人 富士通株式会社
Applicant(s): 株式会社 P F U



2 0 0 3 年 1 0 月 2 4 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康 夫



出証番号 出証特 2 0 0 3 - 3 0 8 8 2 8 1

【書類名】 特許願

【整理番号】 0253075

【提出日】 平成14年12月26日

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 3/06

【発明の名称】 R A I D装置及びその論理デバイス拡張方法

【請求項の数】 10

【発明者】

【住所又は居所】 神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号 富士通株式会社内

【氏名】 伊藤 実希夫

【発明者】

【住所又は居所】 神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号 富士通株式会社内

【氏名】 大黒谷 秀治郎

【発明者】

【住所又は居所】 神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号 富士通株式会社内

【氏名】 池内 和彦

【発明者】

【住所又は居所】 神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号 富士通株式会社内

【氏名】 吉屋 行裕

【発明者】

【住所又は居所】 神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号 富士通株式会社内

【氏名】 大御堂 邦子

【発明者】

【住所又は居所】 石川県河北郡宇ノ気町字宇野気ヌ 9 8 番地の 2 株式会
社ピーエフユー内

【氏名】 平野 忠司

【特許出願人】

【識別番号】 000005223

【氏名又は名称】 富士通株式会社

【特許出願人】

【識別番号】 000136136

【氏名又は名称】 株式会社ピーエフユー

【代理人】

【識別番号】 100094514

【弁理士】

【氏名又は名称】 林 恒徳

【選任した代理人】

【識別番号】 100094525

【弁理士】

【氏名又は名称】 土井 健二

【手数料の表示】

【予納台帳番号】 030708

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9704944

【包括委任状番号】 0215696

【プルーフの要否】 要

【書類名】 明細書**【発明の名称】 R A I D 装置及びその論理デバイス拡張方法****【特許請求の範囲】**

【請求項 1】 R A I D 構成定義に従い、データを分解し、複数の物理ディスク装置に、並列にリード／ライトする R A I D 装置において、

上位装置からの I / O 要求に応じて、前記 R A I D 構成定義による R L U マッピングに従い、前記複数の物理ディスク装置をアクセスする制御部と、

少なくとも旧 R A I D レベルと旧論理デバイス数を定義した旧 R A I D 構成定義情報と、少なくとも新 R A I D レベルと新論理デバイス数を定義した新 R A I D 構成定義情報とを格納するテーブルと、

旧 R A I D 構成から新 R A I D 構成に変更するため、データを一時格納するためのキャッシュメモリとを有し、

前記制御部は、前記テーブルの前記旧 R A I D 構成定義による R L U マッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出し、前記テーブルの前記新 R A I D 構成定義による R L U マッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトする

ことを特徴とする R A I D 装置。

【請求項 2】 前記制御部は、前記旧 R A I D 構成定義の R A I D レベルによる R L U マッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出し、前記新 R A I D 構成定義の R A I D レベルによる R L U マッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトする R A I D レベル変換処理を行う

ことを特徴とする請求項 1 の R A I D 装置。

【請求項 3】 前記制御部は、前記旧 R A I D 構成定義の前記論理デバイス数による R L U マッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出し、前記新 R A I D 構成定義の前記論理デバイス数による R L U マッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトする容量増加処理を行う

ことを特徴とする請求項1のRAID装置。

【請求項4】前記制御部は、前記旧RAID構成から前記新RAID構成への変換をシーケンシャルに実行し、且つその進捗状況を管理するとともに、前記変換中に、前記上位装置からのI/O要求に対し、変換済み領域かを判定し、変換済み領域に対しては、前記新RAID構成定義で、前記I/O要求を実行し、変換済みでない領域に対しては、前記旧RAID構成定義で、前記I/O要求を実行する

ことを特徴とする請求項1のRAID装置。

【請求項5】前記制御部は、新RAID構成定義によるRLBAを上位のLBAに変換した後、前記上位のLBAから前記旧RAID構成定義によるRLUマッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出し、前記RLBAから前記新RAID構成定義によるRLUマッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトする

ことを特徴とする請求項1のRAID装置。

【請求項6】RAID構成定義に従い、データを分解し、複数の物理ディスク装置に、並列にリード/ライトするRAID装置の論理デバイス拡張方法において、

少なくとも旧RAIDレベルと旧論理デバイス数を定義した旧RAID構成定義情報によるRLUマッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出すステップと、

少なくとも新RAIDレベルと新論理デバイス数を定義した新RAID構成定義情報によるRLUマッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトするステップとからなる

ことを特徴とする論理デバイス拡張方法。

【請求項7】前記読出しステップは、前記旧RAID構成定義のRAIDレベルによるRLUマッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出すステップからなり、

前記ライトステップは、前記新RAID構成定義のRAIDレベルによるRL

Uマッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトするステップからなる

ことを特徴とする請求項6の論理デバイス拡張方法。

【請求項8】前記読出しステップは、前記旧RAID構成定義の前記論理デバイス数によるRLUマッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出すステップからなり、

前記ライトステップは、前記新RAID構成定義の前記論理デバイス数によるRLUマッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトする容量増加ステップからなる

ことを特徴とする請求項6の論理デバイス拡張方法。

【請求項9】前記旧RAID構成から前記新RAID構成への変換をシーケンシャルに実行するその進捗状況を管理するステップと、

前記変換中に、前記上位装置からのI/O要求に対し、変換済み領域かを判定するステップと、

変換済み領域に対しては、前記新RAID構成定義で、前記I/O要求を実行するステップと、

変換済みでない領域に対しては、前記旧RAID構成定義で、前記I/O要求を実行するステップとを更に有する

ことを特徴とする請求項6の論理デバイス拡張方法。

【請求項10】前記読出しステップは、新RAID構成定義によるRLBAを上位のLBAに変換した後、前記上位のLBAから前記旧RAID構成定義によるRLUマッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出すステップからなり、

前記ライトステップは、前記RLBAから前記新RAID構成定義によるRLUマッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトするステップからなる

ことを特徴とする請求項6の論理デバイス拡張方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、磁気ディスク等の物理ディスクを使用してデータを冗長管理する R A I D 装置及びその論理デバイス拡張方法に関し、特に、R A I D グループの容量を増加し、又は R A I D グループに冗長性を付加する R A I D 装置及びその論理デバイス拡張方法に関する。

【0002】**【従来の技術】**

磁気ディスク、光磁気ディスク、光ディスク等の記憶媒体を利用したストレージ装置では、データ処理装置の要求で、記憶媒体を実アクセスする。データ処理装置が、大容量のデータを使用する場合には、複数のストレージ装置と制御装置とを備えたストレージシステムを利用する。

【0003】

このようなストレージシステムでは、保存データの信頼性や、装置の信頼性を向上するため、冗長構成を採用している。冗長構成のため、通常、ディスク装置は、R A I D (Redundant Array of Inexpensive (or Independent) Disks) というディスクの多重化構成を採用する。R A I D 機能には、R A I D 0, R A I D 1, R A I D 0+1, R A I D 2, R A I D 3, R A I D 4, R A I D 5 が知られている。

【0004】

このような R A I D 構成は、R A I D レベルは、システムで固定であった。しかし、冗長度を大きくすると、信頼性は向上するが、性能は低下し、又冗長度を小さくすると、信頼性は低下するが、性能は向上する、という相反した特性がある。ユーザーのシステム構成により、冗長度を設定するが、システム導入後、ユーザーが、冗長度を変更したいとの要求がある。冗長度の変更は、システムを停止すれば、容易に冗長度を変更できる。

【0005】

しかし、オンラインシステムを構築した場合に、システムを停止することなく、活性状態で、冗長度を変更することが望ましい。従来、パリティブロックの削減、増加により、冗長度を活性状態で変更する方法が提案されている（例えば、

特許文献1)。

【0006】

この提案では、RAID5の構成において、物理ディスク群からデータをキャッシュメモリに読み出し、2パリティ又は1パリティから1パリティ又は0パリティへの冗長度削減、又は0パリティ又は1パリティから1パリティ又は2パリティへの冗長度増加を行い、キャッシュメモリから物理ディスク群へ書込むものである。

【0007】

この冗長度変換処理中、ホストからI/O要求を受けると、冗長度変換処理を中断し、I/O要求が、冗長度変更済みの領域に対するものか、冗長度変更前の領域に対するものかを判定し、I/O要求を実行するものである。この冗長度削減では、実ディスクの削減を行い、冗長度の増加では、実ディスクの増加を行う必要があった。

【0008】

【特許文献1】

特開平7-306758号公報(図2、図7、図8)

【0009】

【発明が解決しようとする課題】

しかしながら、従来技術は、活性状態での冗長度変更は可能であるものの、パリティブロックの数を単純に、削減、増加する技術であるため、種々のRAIDレベルの変更に対応ができないという問題があり、冗長度の変更範囲が限られていた。

【0010】

更に、活性状態で、RAIDレベルを変更せずに、RAIDグループの容量を増加するという要求を、実現できないという問題があった。

【0011】

従って、本発明の目的は、活性状態でのRAIDレベルの変更範囲を拡大するためのRAID装置及び論理デバイス拡張方法を提供するにある。

【0012】

又、本発明の他の目的は、物理ディスク数の変更なしに、活性状態で R A I D レベルを変更するための R A I D 装置及び論理デバイス拡張方法を提供するにある。

【 0 0 1 3 】

更に、本発明の他の目的は、活性状態で、R A I D レベルを変更せずに、R A I D グループの容量を増大するための R A I D 装置及び論理デバイス拡張方法を提供するにある。

【 0 0 1 4 】

【課題を解決するための手段】

この目的の達成のため、本発明は、R A I D 構成定義に従い、データを分解し、複数の物理ディスク装置に、並列にリード／ライトする R A I D 装置であり、上位装置からの I / O 要求に応じて、前記 R A I D 構成定義による R L U マッピングに従い、前記複数の物理ディスク装置をアクセスする制御部と、少なくとも旧 R A I D レベルと旧論理デバイス数を定義した旧 R A I D 構成定義情報と、少なくとも新 R A I D レベルと新論理デバイス数を定義した新 R A I D 構成定義情報とを格納するテーブルと、旧 R A I D 構成から新 R A I D 構成に変更するため、データを一時格納するためのキャッシュメモリとを有し、前記制御部は、前記テーブルの前記旧 R A I D 構成定義による R L U マッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出し、前記テーブルの前記新 R A I D 構成定義による R L U マッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトする。

【 0 0 1 5 】

又、本発明の論理デバイス拡張方法は、R A I D 構成定義に従い、データを分解し、複数の物理ディスク装置に、並列にリード／ライトする R A I D 装置の論理デバイス拡張方法において、少なくとも旧 R A I D レベルと旧論理デバイス数を定義した旧 R A I D 構成定義情報による R L U マッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出すステップと、少なくとも新 R A I D レベルと新論理デバイス数を定義した新 R A I D 構成定義情報による R L U マッピングに従い、前記キャッシュメモリに読み出したデータを、

前記複数の物理ディスク装置にライトするステップとを有する。

【0016】

本発明では、少なくとも、RAIDレベルと論理デバイス数を定義した新旧のRAID構成定義情報を使用し、それぞれによりRLUマッピングして、RAID構成を変更するため、多様なRAIDレベルの変換、容量増加を実現できる。

【0017】

又、本発明では、好ましくは、前記制御部は、前記旧RAID構成定義のRAIDレベルによるRLUマッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出し、前記新RAID構成定義のRAIDレベルによるRLUマッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトするRAIDレベル変換処理を行う。

【0018】

又、本発明では、好ましくは、前記制御部は、前記旧RAID構成定義の前記論理デバイス数によるRLUマッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出し、前記新RAID構成定義の前記論理デバイス数によるRLUマッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトする容量増加処理を行う。

【0019】

又、本発明では、好ましくは、前記制御部は、前記旧RAID構成から前記新RAID構成への変換をシーケンシャルに実行し、且つその進捗状況を管理するとともに、前記変換中に、前記上位装置からのI/O要求に対し、変換済み領域かを判定し、変換済み領域に対しては、前記新RAID構成定義で、前記I/O要求を実行し、変換済みでない領域に対しては、前記旧RAID構成定義で、前記I/O要求を実行する。

【0020】

又、本発明では、好ましくは、前記制御部は、新RAID構成定義によるRLBAを上位のLBAに変換した後、前記上位のLBAから前記旧RAID構成定義によるRLUマッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出し、前記RLBAから前記新RAID構成定義による

R L Uマッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトする。

【0021】

又、本発明では、好ましくは、前記制御部は、前記旧 R A I D構成から前記新 R A I D構成への変換後、前記旧 R A I D構成定義を前記テーブルから削除する。

【0022】

又、本発明では、好ましくは、前記制御部は、指示された新 R A I D構成定義のパラメータと前記旧 R A I D構成定義とに従い、前記新 R A I D構成定義を、前記テーブルに作成する。

【0023】

又、本発明では、好ましくは、前記制御部は、変換領域に対応する前記キャッシュメモリの領域を獲得した後、前記旧 R A I D構成から前記新 R A I D構成への変換をシーケンシャルに実行する。

【0024】

又、本発明では、好ましくは、前記制御部は、前記変換領域に対応する前記キャッシュメモリの領域が獲得できない時は、前記変換処理を複数回に分けて実行する。

【0025】

又、本発明では、好ましくは、前記制御部は、前記 R A I D構成のストライプに対応したストライプデプス、ストライプサイズに応じて、前記 R L Uマッピングする。

【0026】

【発明の実施の形態】

以下、本発明の実施の形態を、ストレージシステム、R A I D構成、L D E、L D E全体処理、L D E詳細処理、容量増加処理／R A I Dレベル変換処理、他の実施の形態の順で説明する。

【0027】

[ストレージシステム]

図 1 は、本発明の一実施の形態のストレージシステムの構成図であり、磁気ディスクを使用した R A I D (Redundant Arrays of Inexpensive Disk) システムを示す。

【0028】

図 1 に示すように、ストレージシステムは、一対の磁気ディスクコントローラ (以下、コントローラという) 1、2 と、この一対のコントローラ 1、2 にライン 11、12 で接続された多数の磁気ディスク装置 50-1 ~ 50-m、52-1 ~ 52-n とからなる。

【0029】

コントローラ 1、2 は、直接又はネットワーク機器を介し、ホストやサーバーに接続され、ホストやサーバーの大量のデータを、R A I D ディスクドライブ (磁気ディスク装置) へ高速かつ、ランダムに読み書きが出来るシステムである。一対のコントローラ 1、2 は、同一の構成を有し、C A (Channel Adapter) 11、12、21、22 と、C M (Centralized Module) 10、15 ~ 19、20、25 ~ 29 と、D A (Device Adapter) 13、14、23、24 のファンクションモジュールによって構成されている。

【0030】

C A (Channel Adapter) 11、12、21、22 は、ホストを結ぶホスト・インタフェースの制御をつかさどる回路であり、例えば、ファイバーチャネル回路 (F C) と D M A (Direct Memory Access) 回路等で構成される。D A (Device Adapter) 13、14、23、24 は、ディスクデバイス 50-1 ~ 50-m、52-1 ~ 52-m を制御するため、ディスクデバイスとコマンド、データのやり取りを行う回路であり、例えば、ファイバーチャネル回路 (F C) と D M A 回路等で構成される。

【0031】

C M (Centralized Module) は、C P U 10、20 と、ブリッジ回路 17、27 と、メモリー (R A M) 15、25 と、フラッシュメモリ 19、29 と、I O ブリッジ回路 18、28 とを有する。メモリー 15、25 は、バッテリーでバックアップされ、その一部が、キャッシュメモリ 16、26 として使用される。

【0032】

CPU10, 20は、ブリッジ回路17を介し、メモリー15, 25、フラッシュメモリ19, 29、IOブリッジ回路18, 28に接続される。このメモリー15, 25は、CPU10, 20のワーク領域に使用され、フラッシュメモリ19, 29は、CPU10, 20が実行するプログラムを格納する。

【0033】

このプログラムとして、OS, BIOS (Basic Input/Output System), ファイルアクセスプログラム (リード/ライトプログラム)、RAID管理プログラム等の制御プログラム (モジュール) を格納する。CPU10, 20は、このプログラムを実行し、後述するように、リード/ライト処理、RAID管理処理等を実行する。

【0034】

PCI (Personal Computer Interface) バス31は、CA11, 12, 21, 22と、DA13, 14, 23, 24とを接続するとともに、IOブリッジ回路18を介し、CPU10, 20、メモリー15, 25を接続する。更に、PCIバス31には、PCI-ノードリンクブリッジ回路30, 40が接続される。コントローラ1のPCI-ノードリンクブリッジ回路30は、コントローラ2のPCI-ノードリンクブリッジ回路40と接続され、コントローラ1, 2間のコマンド、データの通信を行う。

【0035】

図1では、例えば、コントローラ1は、ディスク装置50-1~50-mを担当し、コントローラ2は、ディスク装置52-1~52-nを担当する。又、ディスク装置50-1~50-mと、52-1~52-nとが、RAID5の構成を有する。

【0036】

キャッシュメモリ16, 26は、各々、担当するディスク装置のデータの一部を格納し、ホストからのライトデータを格納する。CPU10, 20は、CA11, 12, 21, 22を介しホストからのリード要求を受けて、キャッシュメモリ16, 26を参照し、物理ディスクへのアクセスが必要かを判定し、必要であ

れば、ディスクアクセス要求をDA13, 14, 23, 24に要求する。又、CPU10, 20は、ホストからのライト要求を受けて、ライトデータをキャッシュメモリ16, 26に書込み、且つ内部でスケジュールされるライトバック等をDA13, 14, 23, 24に要求する。

【0037】

[RAID構成]

図2は、CPU10が実行するタスクと制御モジュールの構成図である。図2に示すように、CM-CAドライバ32は、CA11, 12をドライブするドライバである。Basic Task33は、CPU10が実行する基本タスクであり、資源管理を行うリソーススレッド（リソース制御モジュール）35、コピー処理を行うコピースレッド（コピー制御モジュール）36、キャッシュメモリ16の制御を行うキャッシュスレッド（キャッシュメモリ制御モジュール）37、RAID構成制御を行うRAIDスレッド（RAID制御モジュール）39、メンテナンスエージェント（メンテナンス制御モジュール）34、OVSMスレッド（OVSM制御モジュール）38とを有する。

【0038】

OVSMスレッド38は、後述するように、クイックフォーマット（QF）や論理デバイス拡張（LDE）のスケジューリング、RAIDスレッド39への論理フォーマット（LF）依頼、LDE進捗管理を行う。メンテナンスエージェント34は、OVSMスレッド38への各種の通知を行う。

【0039】

CM-DAドライバ42は、CM13, 14をドライブするドライバである。CM-CMドライバ43は、前述のPCI-ノードブリッジ回路30をドライブして、CM間の通信を行う。

【0040】

保守PC（パーソナルコンピュータ）3は、図1では、図示していないが、ブリッジ回路17のLANポートに接続され、ストレージシステムの保守を行うものであり、本発明では、LDEの要求、LDE進捗報告の表示等を行う。HTTPドライバ44は、保守PC3とのHTTP（Hyper Text Transfer Protocol）

のドライバである。

【0041】

C G I (Computer Graphic Interface) タスク45は、保守PC3からのLDE要求の受け付けと、保守PC3へのLDE進捗報告を行う。メンテナンスドライバ46は、メンテナンスのためのメンテナンスPCスレッド47とシステム制御スレッド48のドライバである。

【0042】

メンテナンスPCスレッド(制御モジュール)47は、保守PC3からの要求に応じ、システム制御スレッド48を起動する。システム制御スレッド(制御モジュール)48は、QF/LDE開始時のRLU(Raid Logical Unit)テーブル(図3、図14で後述する)の変更、OVSMスレッド38へのQF/LDE開始指示、QF/LDE終了時のRLUテーブル(図3、図14にて後述する)の変更を行う。

【0043】

本発明では、後述するように、保守PC3のLDE要求により、OVSMスレッド38が、キャッシュスレッド37、RAIDスレッド39を制御し、LDE処理を実行する。

【0044】

図3は、RAID論理空間の説明図、図4は、RAID5の説明図、図5は、RAID5のアドレス計算フロー図、図6は、RAID0+1の説明図、図7は、RAID0+1のアドレス計算フロー図である。

【0045】

図3に示すように、ホストから見たRAID論理空間は、ホストの論理空間であるOLU(ホスト論理ユニット)と、RAIDグループの論理空間であるRLU(RAID論理ユニット)と、RAIDグループを構成するデバイスの論理空間であるDLU(デバイス論理ユニット)と、物理ディスクの論理空間であるPLU(物理論理ユニット)との層構造で示される。

【0046】

RAID構成では、RAID論理空間は、OLUテーブル70の開始RLBA

(RAID論理ブロックアドレス)で、OLUと関連づけられ、RAID空間は、RLUテーブル72で定義される。RLUテーブル72は、RAIDレベル、構成ディスク数、RAIDストライプデプス、RAIDストライプサイズ、対応DLU番号を格納する。

【0047】

DLU空間は、DLUテーブル74で定義される。DLUテーブル74は、構成ディスク数、RAIDストライプデプス、RAIDストライプサイズ、対応PLU番号を格納する。DLU空間及びDLUテーブル74は、ミラーリングで使用される。PLU空間は、PLUテーブル76で定義される。PLUテーブル76は、開始PLBA (物理論理ブロックアドレス) を格納する。

【0048】

具体的に説明する。図4に示すように、RAID5 (3+1) の場合には、RLU=DLUであり、RLUテーブル72は、RAIDレベル=RAID5, 構成ディスク数=4、対応DLU番号=PLU番号 (0~3) となる。又、RAID空間は、構成ディスクで、ストライピングされており、RAID空間を、構成ディスク番号とストライプ番号とでマッピングする。この升目をストリップ (Strip) といい、Strip番号が付与される。このStripのサイズは、Strip Depth (又はStripe Depth) で定義され、1ストライプのサイズは、Stripe Sizeで定義される。

【0049】

従って、図5で説明するように、R (RAIDグループ) LBAは、構成ディスク数、Strip Depth、Stripe Sizeで、PLULBA, メンバーディスクの順番に変換できる。

【0050】

(S10) ホストLBA (論理ブロックアドレス) を、ホストLBAにOLUテーブル70の開始RLBAを加算して、RLBAを求める。

【0051】

(S12) ストリップ内のブロックカウントを、RLULBA/Strip Size (Stripe Depth) の余りで計算する。

【0052】

(S14) Strip番号を、RLULBA/Strip Depthから計算する。

【0053】

(S16) Stripe番号を、RLULBA/Stripe Sizeから計算する。

【0054】

(S18) メンバーディスクの順番を、Strip番号/メンバーディスクの数の余りから計算する。

【0055】

(S20) 物理ディスク(PLU)のLBAを、(Stripe番号×Strip Size) + ストリップ内のブロックカウントから計算する。

【0056】

これにより、メンバーディスクの順番(PLU番号)と、PLULBAとから、PLUテーブル76を用いて、実ブロックアドレスを計算する。

【0057】

同様に、図6に示すように、RAID0+1(4+4)の場合には、RLU≠DLUであり、RLUテーブル72は、RAIDレベル=RAID0+1, 構成ディスク数=4(DLU)、対応DLU番号=0~3となる。又、RAID空間は、DLU構成ディスクで、ストライピングされており、RAID空間を、DLU構成ディスク番号とストライプ番号とでマッピングする。この升目をストリップ(Strip)といい、Strip番号が付与される。このStripのサイズは、Strip Depth(又はStripe Depth)で定義され、1ストライプのサイズは、Stripe Sizeで定義される。

【0058】

従って、図7で説明するように、R(RAIDグループ)LBAは、構成ディスク数、Strip Depth、Stripe Sizeで、PLULBA, メンバーディスクの順番に変換できる。

【0059】

(S22) ホストLBA (論理ブロックアドレス) を、ホストLBAにOLU
テーブル70の開始RLBAを加算して、RLBAを求める。

【0060】

(S24) ストリップ内のブロックカウントを、 $RLULBA / \text{Stripe Depth}$ の余りで計算する。

【0061】

(S26) Stripe番号を、 $RLULBA / \text{Stripe Size}$ から計算する。

【0062】

(S28) メンバーディスクの順番を、 $(RLULBA / \text{Stripe Size} \text{ の余り}) / \text{Strip Size}$ から計算する。

【0063】

(S30) 物理ディスク (PLU) のLBA (=DLUのディスク) を、 $(\text{Stripe 番号} \times \text{Strip Size}) + \text{ストリップ内のブロックカウント}$ から計算する。

【0064】

これにより、メンバーディスクの順番 (DLU番号) と、PLULBAとから、PLUテーブル76を用いて、実ブロックアドレスを計算する。

【0065】

[LDE]

次に、LDEを説明する。Logical Device Expansion(LDE)は、(1) ディスク装置を増設したり、RAIDレベルを変換することで、RAIDグループの容量を増加させる、(2) RAIDレベルを変換することで、RAIDグループに冗長性を付加する、機能である。RAID容量を拡張する方法として、RAIDグループへの新規ディスク装置増設と、RAIDレベル変換とがある。

【0066】

図8及び図9は、RAIDグループへの新規ディスク装置増設 (容量増加) の説明図である。容量増加は、RAIDグループにユーザデータを保持したまま、新規ディスク装置(New Disk)を追加して、RAIDグループの容量を増やす機能

である。

【0067】

例えば、図8に示すように、RAID5 (4+1) に、新規ディスクを1台追加して、RAID5 (5+1) にする。追加ディスクの容量が、36GBなら、RAIDグループ144GBを、180GBに容量を増やす。

【0068】

また、図9に示すように、全ユーザデータを、容量の大きな新規ディスク装置で構成されたRAIDグループに遷移することでも、RAID容量の拡張を実現できる。例えば、18GBのディスク装置で構成されたRAID5 (4+1) 容量72GBのユーザデータを、新規ディスク装置 (36GB) 5台を用いて構成されたRAID5 (4+1) 容量144GBへ遷移し、RAID容量を増やす。

【0069】

次に、RAIDレベル変換を、図10乃至図12は、RAIDレベル変換の説明図である。RAIDレベル変換は、ユーザデータをRAIDグループに保持したまま、RAIDレベルの変更を行う機能である。

【0070】

例えば、図10に示すように、複数台 (4台) の新規ディスク装置を追加し、RAID5 (4+1) からRAID0+1 (4+4) への変更を行う。

【0071】

また、RAIDグループを構成している既存ディスク装置を用いず、新規ディスク装置だけを用いてのRAIDレベルの変更も出来る。例えば、図11に示すように、18GBディスク装置4台構成のRAID5 (3+1) を、73GBディスク装置2台構成のRAID1へ変更する。

【0072】

本実施の形態では、可能なRAIDレベル変換は、図12に示すように、RAID0からRAID0+1、RAID1、RAID5、RAID0+1とRAID1、RAID5と相互間、RAID1とRAID5との相互間である。RAID0への変換は、冗長性を失うことになるため実施しない。

【0073】

このLDE実行中、LDEを担当するCMを切り替えることができる。又、LDEを、パワーオフ/オンを跨いで継続され、停電/復電を跨いで継続される。更に、CM活性増設・活性交換は実行できる。

【0074】

但し、RAID1へのRAIDレベル変換は、既存ディスクを使用せずに新規ディスクのみで作成し、その新規ディスクは、既存ディスクよりも容量が大きくないことが必要である。又、RAIDレベル変換について、RAID0への変換と容量が減少する変換は行わない。

【0075】

[LDE全体処理]

次に、図13及び図14により、図2の構成のLDE時のCGIタスク45からの流れを説明する。先ず、CGIタスク45より保守タスク47へLDE設定通知を行う。このときのパラメタは、図9で説明するように、LDE内容（追加、レベル変換等）、RLUN、RAIDレベル、新規構成ディスクである。保守タスク47は、装置状態がLDE実行可能かどうかをチェックする。更に、保守タスク47は、システム制御モジュール48に対して、設定パラメタがLDE実行可能な条件（図12参照）を満たしているかどうかのチェックを要求する。

【0076】

CGIタスク45は、保守タスク47からLDE実行可能な応答を受けて、保守タスク47へLDE設定を通知する。これに応じて、保守タスク47は、システム制御48にLDE設定通知を行う。システム制御48は、キャッシュ制御37に対して、WTH (Write Through) で動作するよう通知する。キャッシュ制御37は、以降、WTHで動作する。

【0077】

次に、保守タスク47は、構成変更を行うためシステム制御48に対して、Suspendを通知する。システム制御48は、それを受けて保守エージェント34へSuspend要求を出す。保守エージェント34は、Backend (Cacheモジュール37、RAIDモジュール39) に対してSuspend要求を出す。これにより、対象RLUを含めてすべてのI/Oを一時的に抑制する。

【0078】

更に、システム制御48は、Suspend処理が完了すると、OVSM38に対して構成変更を要求する。OVSM38は、LDEを開始するための構成を作成する。図14に示すように、現在のRAID構成定義、即ち、RLU72, DLU74, PLU76を、旧構成のテンポラリRAID定義80にコピーし、前述のCGI45から通知されたLDE内容（追加、レベル変換等）、RLUN, RAIDレベル、新規構成ディスクから新構成のRAID定義82、即ち、RLU72, DLU74, PLU76を作成する。OVSM38は、変更内容を各モジュールに対して配信する。

【0079】

この処理を終了後、システム制御48は、保守エージェント34に対して、LDE実行要求を出す。それを受けて保守エージェント34は、OVSM38に対して、LDE実行要求を出す。OVSM38は、LDE初期処理を実行し、保守エージェント34へ応答を返す。その後、OVSM38は、図14のLDE進捗状況を作成し、LDE処理を実行する。保守エージェント34は、応答をシステム制御48へ返す。

【0080】

この後、システム制御48は、保守エージェント34に対して、Resumeを通知する。保守エージェント34は、BackendへResume要求を出す。これによって、I/Oを再開する。システム制御48は、キャッシュ制御37に対して、WB(Write Back)modeへ戻るように通知する。システム制御37は、CGIタスク45に対して、LDEが開始されたことを通知する。

【0081】

次に、CGIタスク45は、保守タスク47経由で、OVSM38に対して、LDE進捗情報獲得要求を出し、進捗情報を獲得する。OVSM38は、LDE処理が完了すると、CVMmodeのLDEFlagをOFFにする等の構成変更を行い、旧構成情報を削除する。OVSM38は、Backendに対して、構成を配信する。OVSM38は、LDE処理の後処理を行う。

【0082】

[LDE 詳細処理]

次に、LDE 処理を、図 15 乃至図 23 で説明する。図 15 は、LDE 制御モジュールの構成図、図 16 は、図 15 の LDE 処理の説明図、図 17 は、図 15 の LDE 起動時の処理フロー図、図 18 は、図 15 のキャッシュメモリ獲得処理フロー図、図 19 は、図 15 のシーケンシャル LDE 処理の説明図、図 20 は、図 15 の進捗状況管理による I/O 競合時の処理フロー図である。

【0083】

図 15 に示すように、OVSM38 は、Expansion Block を決定し、キャッシュメモリ 16 の領域を獲得する処理 90 を実行し、RAID スレッド 39 に、LDE のリード／ライトを依頼する。又、RAID スレッド 39 の LDE の進捗状況を管理する処理 92 を行う。

【0084】

RAID スレッド 39 は、キャッシュスレッド 37 と共同して、Expansion 実施前の RAID グループ構成（図 14 の旧構成テーブル 80）で、ディスク装置 50 へのリード処理を行い、キャッシュメモリ 16 の獲得領域にステージング処理 94 を行う。

【0085】

次に、RAID スレッド 39 は、Expansion 実施後の RAID グループ構成（図 14 の新構成テーブル 82）で、キャッシュメモリ 16 からディスク装置 50 へのライト（ライトバック）処理 96 を行う。

【0086】

この時、LDE では、リード／ライト処理時のアドレスが異なる。この処理を、増設の場合には、新ディスク装置を含めた領域まで実施したところで、Expansion が完了する。この LDE の実施にあたり、図 14 で説明したように、Expansion 前（旧構成）の RAID グループ構成情報は、テンポラリ RAID グループ定義情報 80 として引継ぎ、Expansion 処理の終了時に削除する。

【0087】

図 16 は、RAID における Expansion 処理（容量拡張処理）の説明図であり、4 台のディスク装置に新規ディスク装置 1 台を追加する例を示す。新構成のメ

ンバーディスク装置数（ここでは、5台）分のストライプ（図の枠内）を、旧構成のディスク装置（4台）からリードする。次に、旧構成のメンバーディスク装置数分のストライプを、新構成のディスク装置5台にライトする。図16では、RAID0+1において、（4+4）構成から、（5+5）構成に変更する例を示し、旧構成（4+4）から5つのストライプをリードし、新構成（5+5）に4つのストライプをライトする。

【0088】

次に、図17により、保守エージェントよりLDE開始要求を受けてからのOVSMの制御シーケンスフローを説明する。保守エージェント34は、OVSM38に対し、LDEの起動を指示する。このときのパラメタは、LDE内容（追加、レベル変換等）、RLUN、RAIDレベル、新規構成ディスクである。

【0089】

OVSM38は、受け取ったパラメタから、構成チェックを行う。即ち、RLUN82とT-RLUN80の比較、LDEタイプ等である。そして、OVSM38は、シーケンシャルExpansion用のACB（Access Control Block）を獲得する。尚、ACBの数だけ平行動作が可能となる。且つExpansionを行うにあたり、ディスク装置上のデータを移動するために、キャッシュメモリ16を獲得する（図18にて詳細に説明する）。

【0090】

更に、OVSM38は、他系CM（図1のコントローラ2）へ制御テーブル及びキャッシュ域獲得を要求する。他系CMは、制御テーブル及びキャッシュ域の獲得を行う。これにより、制御テーブル及びキャッシュ域の二重化が可能となる。OVSM38は、他CMからキャッシュ域獲得の応答を受け取ると、保守エージェント34へLDE起動応答を通知する。

【0091】

そして、OVSM38は、シーケンシャルなExpansion開始をRAIDスレッド39に通知する。即ち、OVSM38は、一回のExpansion処理領域とI/Oの排他を行う（図20で詳述する）。そして、RAIDスレッド39に、Expansion実施中領域について、テンポラリRLUN(Expansion前構成)80でのリ

ード要求を行う。

【0092】

OVSM38は、他系CMとLDEステータス情報の二重化を行う。そして、OVSM38は、RAIDスレッド39に、Expansion実施中領域について、RLUN(Expansion後構成)82でライト要求を行う。OVSM38は、Expansion実施中領域の進捗情報を更新して、二重化されている管理テーブルの進捗情報更新処理を行うために、他系CMへ通信を実施する。OVSM38は、一回のExpansion処理領域とI/Oとの排他を解除する。

【0093】

次に、図15、図17で説明したキャッシュメモリ獲得処理90を、図18及び図19で説明する。図15で説明したように、Expansion処理中は、ディスク上のデータを移動させるので、一時的にキャッシュメモリ16にデータを置く。そのため、Sequential Expansion実施前にキャッシュメモリ16を獲得しておく。キャッシュメモリ16は、一度に獲得できる大きさが決まっているので、必要容量を獲得するためには、ACBを必要数獲得してから実施する。

【0094】

図18に示すように、1回の処理サイズを決定し、その処理サイズが、一度に獲得できるキャッシュメモリの限界サイズかを判定する。獲得可能なサイズなら、キャッシュスレッド37に1回だけ、獲得要求を発する。一方、獲得可能でないなら、複数回獲得できるかを判定し、複数回獲得できれば、必要回数要求し、複数回獲得できなければ、1回だけ要求し、複数回使いまわす。即ち、Expansion実施領域のLDEを、一度に実施せずに複数回に分けて実施する。または、必要ACB等を獲得してから、平行処理を行うことにする。

【0095】

例えば、図18のようなExpansion処理において、一回のExpansion領域を複数回のリード／ライトを実施していき、次の領域へ進む。具体的に説明する。

【0096】

(1)リード処理をExpansion後のストライプ分(3ストライプ)行う。

【0097】

(2) (1)で読み込んだ領域について、ライト処理を行う。

【0098】

(3) リード処理をExpansion後のストライプ分(1ストライプ)行う。

【0099】

(4) (3)で読み込んだ領域について、ライト処理を行う。

【0100】

図18では、RAID0+1の(4+4)からRAID0+1の(5+5)への構成変更の例を示し、Expansionブロック数=旧構成MemberDisk数×新構成MemberDisk数=4×5=20ブロック(ストリップ)であり、ロックするストライプ数=キャッシュメモリサイズ÷(新構成MemberDisk数×60KB)=1MB÷(5×60KB)=3ストライプとなる。又、Expansion実行数=Expansionブロック数÷(ロックするストライプ数×新構成MemberDisk数)=20÷(3×5)=2回であり、最後のExpansion実行ブロック数=Expansionブロック数%・(ロックするストライプ数×新構成MemberDisk数)=20%・(3×5)=5ブロックとなる。但し、キャッシュ域は、1MB、Stripは、60KBとする。

【0101】

又、RAID構成の(14+1)から(15+1)へのExpansionの場合には、Expansionブロック数=14×15=210ブロック、ロックするストライプ数=1MB÷(15×60KB)=1ストライプ、Expansion実行数=210÷(1×15)=14回、最後のExpansion実行ブロック数=210%・(1×15)=15ブロックとなる。

【0102】

次に、図20により、OVSM38のI/O競合処理を説明する。

【0103】

(S40) OVSM38は、RLUを、RLBAにより、Expansion実施済みか、実施中か、未実施かを判定する進捗管理機能92(図15参照)により、ホストI/O要求のRLBAが、Expansion実施領域か未実施領域かの判断を行う。実施済み領域の場合、RAIDスレッド39へ新構成情報でI/O要求を行い、ホストI/O要求を実行する。

【0104】

(S 4 2) 未実施の場合には、Expansion実施中領域かの判断を行う。実施中の場合には、Expansion処理が終了するまで、ホスト I/O 要求を待たせ、終了後に、RAIDスレッド39へI/O要求を行う。一方、実施中領域でない場合には、Expansion未実施領域のため、RLU構成情報をテンポラリRLU構成情報(旧構成)80に切り替えてから、RAIDスレッド39へI/O要求を行う。

【0105】

次に、図15で説明したリード処理、ライト処理を、図21乃至図23で説明する。図21は、リード処理のリードアドレス生成処理フロー図である。

【0106】

(S 5 0) 先ず、新RAID構成定義82で、新RLBA (RLUBLA) を生成する。

【0107】

(S 5 2) 図3で説明したOLUテーブル70を使用して、OLBAに逆変換する。

【0108】

(S 5 4) OLBAを旧構成のOLUテーブル70で、旧RLBAに変換する。

【0109】

(S 5 6) 旧RLBAを、旧RAID構成定義のRLUテーブル72、DLUテーブル74を使用して、旧DLBAに変換する。

【0110】

(S 5 8) 旧RAID構成定義のPLUテーブル76を使用して、旧DLBAをPLBAに変換し、リードブロックアドレスを得る。

【0111】

このリードブロックアドレスで、ディスク装置をリードし、データをキャッシュメモリ16にリード(ステージング)する。

【0112】

図22は、ライト処理のライトアドレス生成処理フロー図である。

【0113】

(S60) 先ず、新RAID構成定義82で、新RLBA (RLUBLA) を生成する。

【0114】

(S62) 図3で説明した新RAID構成定義のRLUテーブル72、DLUテーブル74を使用して、新DLBAに変換する。

【0115】

(S64) 新RAID構成定義のPLUテーブル76を使用して、新DLBAをPLBAに変換し、ライトブロックアドレスを得る。

【0116】

このライトブロックアドレスで、キャッシュメモリ16のデータを、ディスク装置にライトバックする。

【0117】

図23は、この関係を示す説明図であり、新RAID構成定義での新RLBAを上位のホストLBAに直し、このホストLBAから旧RAID定義80で、旧RAID定義でのリードPLBAに変換し、リードする。又、新PLBAを、新RAID構成定義82で、新RAID定義でのライトPLBAに変換し、リードしたデータをライトする。

【0118】

このようにすると、図4乃至図7で説明した既存のPLBA, Strip Size, Strip Depth, Stripe Size, メンバーディスク数を用いたRAIDマッピング処理を利用して、LDE処理を実行できる。又、増設処理とRAIDレベル変換処理を同じ処理で実現できる。

【0119】

[容量増加処理／RAIDレベル変換処理]

図24は、容量増加時の旧RAID定義と新RAID定義の説明図、図25は、その動作説明図である。図24に示すように、旧(テンポラリ)RAID定義80では、RAIDレベル5で、3+1 (PLUNs 10-13) で、メンバーディスク数4、ブロック数400を定義する。一方、新RAID定義82では、RAIDレベル5で、4+1 (PLUNs 10-14) で、メンバーディスク数

5、ブロック数500を定義する。

【0120】

前述の図4及び図5の説明のように、メンバーディスク数の増加で、マッピングが変更され、図25に示すように、RAID5(3+1)が、RAID5(4+1)に増設される。

【0121】

図26は、RAIDレベル変換時の旧RAID定義と新RAID定義の説明図、図27は、その動作説明図、図28は、RLUテーブルとDLUテーブルの遷移の説明図である。

【0122】

図26に示すように、旧(テンポラリ)RAID定義80では、RAIDレベル0+1で、メンバーディスク数4(PLUNs0-3)で、ブロック数400を定義する。一方、新RAID定義82では、RAIDレベル5で、3+1(PLUNs0-3)で、メンバーディスク数4、ブロック数400を定義する。

【0123】

前述の図6及び図7の説明のように、RAID0+1のストライプをリードし、図4及び図5の説明のように、RAID5の定義で、マッピングが変更され、図27に示すように、RAID0+1(4+4)が、RAID5(3+1)にレベル変換される。これにより、図28に示すように、RLUテーブル72、DLUテーブル74が、RAID0+1からRAID5の定義に変更される。

【0124】

[他の実施の形態]

図29は、本発明の他のアドレス変換の説明図である。この実施の形態では、予め、変換の対象となるRAID構成定義間のRLBAの差分をテーブルに求めておき、旧RAID定義のRLBAで、旧RAID定義のリードPLBAを計算する。

【0125】

この旧RAID定義のRLBAを、RLBA間の差分を足し、新RLBAを求め、新RAID定義82で、ライトPLBAを計算する。このようにすると、R

L B A間の差分計算の手間はかかるが、高速にマッピング変換が可能となる。

【0126】

前述の実施の形態では、2 C M（コントローラ）－4 D E（デバイスエンクロージャ）の構成で説明したが、4 台のコントローラを有する4 C M-1 6 D Eの構成でも、各C Mは、保守P C 3のL D E起動により、同様に、L D E処理できる。

【0127】

又、前述の実施の形態では、図12のようなR A I Dレベルで説明したが、これ以外のR A I Dレベル（R A I D 2，R A I D 3，R A I D 4）のストレージシステムに適用できる。又、物理ディスクは、磁気ディスク、光ディスク、光磁気ディスク、各種のストレージデバイスを適用できる。

【0128】

以上、本発明を実施の形態により説明したが、本発明の趣旨の範囲内において、本発明は、種々の変形が可能であり、本発明の範囲からこれらを排除するものではない。

【0129】

（付記1）R A I D構成定義に従い、データを分解し、複数の物理ディスク装置に、並列にリード／ライトするR A I D装置において、上位装置からのI／O要求に応じて、前記R A I D構成定義によるR L Uマッピングに従い、前記複数の物理ディスク装置をアクセスする制御部と、少なくとも旧R A I Dレベルと旧論理デバイス数を定義した旧R A I D構成定義情報と、少なくとも新R A I Dレベルと新論理デバイス数を定義した新R A I D構成定義情報とを格納するテーブルと、旧R A I D構成から新R A I D構成に変更するため、データを一時格納するためのキャッシュメモリとを有し、前記制御部は、前記テーブルの前記旧R A I D構成定義によるR L Uマッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出し、前記テーブルの前記新R A I D構成定義によるR L Uマッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトすることを特徴とするR A I D装置。

【0130】

(付記 2) 前記制御部は、前記旧 R A I D 構成定義の R A I D レベルによる R L U マッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出し、前記新 R A I D 構成定義の R A I D レベルによる R L U マッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトする R A I D レベル変換処理を行うことを特徴とする付記 1 の R A I D 装置。

【0131】

(付記 3) 前記制御部は、前記旧 R A I D 構成定義の前記論理デバイス数による R L U マッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出し、前記新 R A I D 構成定義の前記論理デバイス数による R L U マッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトする容量増加処理を行うことを特徴とする付記 1 の R A I D 装置。

【0132】

(付記 4) 前記制御部は、前記旧 R A I D 構成から前記新 R A I D 構成への変換をシーケンシャルに実行し、且つその進捗状況を管理するとともに、前記変換中に、前記上位装置からの I / O 要求に対し、変換済み領域かを判定し、変換済み領域に対しては、前記新 R A I D 構成定義で、前記 I / O 要求を実行し、変換済みでない領域に対しては、前記旧 R A I D 構成定義で、前記 I / O 要求を実行することを特徴とする付記 1 の R A I D 装置。

【0133】

(付記 5) 前記制御部は、新 R A I D 構成定義による R L B A を上位の L B A に変換した後、前記上位の L B A から前記旧 R A I D 構成定義による R L U マッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出し、前記 R L B A から前記新 R A I D 構成定義による R L U マッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトすることを特徴とする付記 1 の R A I D 装置。

【0134】

(付記 6) 前記制御部は、前記旧 R A I D 構成から前記新 R A I D 構成への変

換後、前記旧 R A I D 構成定義を前記テーブルから削除することを特徴とする付記 1 の R A I D 装置。

【0135】

(付記 7) 前記制御部は、指示された新 R A I D 構成定義のパラメータと前記旧 R A I D 構成定義とに従い、前記新 R A I D 構成定義を、前記テーブルに作成することを特徴とする付記 1 の R A I D 装置。

【0136】

(付記 8) 前記制御部は、変換領域に対応する前記キャッシュメモリの領域を獲得した後、前記旧 R A I D 構成から前記新 R A I D 構成への変換をシーケンシャルに実行することを特徴とする付記 1 の R A I D 装置。

【0137】

(付記 9) 前記制御部は、前記変換領域に対応する前記キャッシュメモリの領域が獲得できない時は、前記変換処理を複数回に分けて実行することを特徴とする付記 8 の R A I D 装置。

【0138】

(付記 10) 前記制御部は、前記 R A I D 構成のストライプに対応したストリップデプス、ストライプサイズに応じて、前記 R L U マッピングすることを特徴とする付記 1 の R A I D 装置。

【0139】

(付記 11) R A I D 構成定義に従い、データを分解し、複数の物理ディスク装置に、並列にリード／ライトする R A I D 装置の論理デバイス拡張方法において、少なくとも旧 R A I D レベルと旧論理デバイス数を定義した旧 R A I D 構成定義情報による R L U マッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出すステップと、少なくとも新 R A I D レベルと新論理デバイス数を定義した新 R A I D 構成定義情報による R L U マッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトすることを特徴とする論理デバイス拡張方法。

【0140】

(付記 12) 前記読出しステップは、前記旧 R A I D 構成定義の R A I D レベ

ルによる R L U マッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出すステップからなり、前記ライトステップは、前記新 R A I D 構成定義の R A I D レベルによる R L U マッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトするステップからなることを特徴とする付記 11 の論理デバイス拡張方法。

【0141】

(付記 13) 前記読出しステップは、前記旧 R A I D 構成定義の前記論理デバイス数による R L U マッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出すステップからなり、前記ライトステップは、前記新 R A I D 構成定義の前記論理デバイス数による R L U マッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトする容量増加ステップからなることを特徴とする付記 11 の論理デバイス拡張方法。

【0142】

(付記 14) 前記旧 R A I D 構成から前記新 R A I D 構成への変換をシーケンシャルに実行するその進捗状況を管理するステップと、前記変換中に、前記上位装置からの I / O 要求に対し、変換済み領域かを判定するステップと、変換済み領域に対しては、前記新 R A I D 構成定義で、前記 I / O 要求を実行するステップと、変換済みでない領域に対しては、前記旧 R A I D 構成定義で、前記 I / O 要求を実行するステップとからなることを特徴とする付記 11 の論理デバイス拡張方法。

【0143】

(付記 15) 前記読出しステップは、新 R A I D 構成定義による R L B A を上位の L B A に変換した後、前記上位の L B A から前記旧 R A I D 構成定義による R L U マッピングに従い、前記複数の物理ディスク装置からデータを前記キャッシュメモリに読出すステップからなり、前記ライトステップは、前記 R L B A から前記新 R A I D 構成定義による R L U マッピングに従い、前記キャッシュメモリに読み出したデータを、前記複数の物理ディスク装置にライトするステップからなることを特徴とする付記 11 の論理デバイス拡張方法。

【0144】

(付記16) 前記旧RAID構成から前記新RAID構成への変換後、前記旧RAID構成定義を削除するステップを更に有することを特徴とする付記11の論理デバイス拡張方法。

【0145】

(付記17) 指示された新RAID構成定義のパラメータと前記旧RAID構成定義とに従い、前記新RAID構成定義を、テーブルに作成するステップを更に有することを特徴とする付記11の論理デバイス拡張方法。

【0146】

(付記18) 変換領域に対応する前記キャッシュメモリの領域を獲得した後、前記旧RAID構成から前記新RAID構成への変換をシーケンシャルに実行するステップを更に有することを特徴とする付記11の論理デバイス拡張方法。

【0147】

(付記19) 前記変換領域に対応する前記キャッシュメモリの領域が獲得できない時は、前記変換処理を複数回に分けて実行するステップを更に有することを特徴とする付記18の論理デバイス拡張方法。

【0148】

(付記20) 前記RAID構成のストライプに対応したストリップデプス、ストライプサイズに応じて、前記RLUマッピングするステップを更に有することを特徴とする付記11の論理デバイス拡張方法。

【0149】**【発明の効果】**

このように、本発明では、少なくとも、RAIDレベルと論理デバイス数を定義した新旧のRAID構成定義情報を使用し、それぞれによりRLUマッピングして、RAID構成を変更するため、多様なRAIDレベルの変換、容量増加を活性状態で実現できる。

【図面の簡単な説明】**【図1】**

本発明の一実施の形態のストレージシステムの構成図である。

【図 2】

図 1 の制御モジュールの構成図である。

【図 3】

図 2 の R A I D 論理空間と論理テーブルの説明図である。

【図 4】

図 2 の R A I D 5 の論理マッピング図である。

【図 5】

図 4 の R A I D 5 の P L B A 計算処理フロー図である。

【図 6】

図 2 の R A I D 0 + 1 の論理マッピング図である。

【図 7】

図 6 の R A I D 0 + 1 の P L B A 計算処理フロー図である。

【図 8】

図 1 の R A I D グループ容量増加の一実施の形態の説明図である。

【図 9】

図 1 の R A I D グループ容量増加の他の実施の形態の説明図である。

【図 1 0】

図 1 の R A I D レベル変換の一実施の形態の説明図である。

【図 1 1】

図 1 の R A I D レベル変換の他の実施の形態の説明図である。

【図 1 2】

図 1 の R A I D レベル変換の関係図である。

【図 1 3】

図 2 の L D E 全体処理フロー図である。

【図 1 4】

図 1 3 の R A I D 構成定義の説明図である。

【図 1 5】

図 2 の L D E 制御モジュールの構成図である。

【図 1 6】

図 15 の L D E マッピング処理の説明図である。

【図 17】

図 15 の L D E 詳細処理フロー図である。

【図 18】

図 17 のキャッシュメモリ獲得処理フロー図である。

【図 19】

図 17 のシーケンシャル L D E の説明図である。

【図 20】

図 15 の I / O 競合処理フロー図である。

【図 21】

図 15 のリードアドレス計算処理フロー図である。

【図 22】

図 15 のライトアドレス計算処理フロー図である。

【図 23】

図 21 及び図 22 のアドレス変換処理の説明図である。

【図 24】

図 8 の容量増加の新旧 R A I D 定義の説明図である。

【図 25】

図 24 の新旧 R A I D 定義による容量増加の説明図である。

【図 26】

図 10 の R A I D レベル変換の新旧 R A I D 定義の説明図である。

【図 27】

図 26 の新旧 R A I D 定義による R A I D レベル変換の説明図である。

【図 28】

図 27 の R A I D レベル変換による R L U テーブル、D L U テーブルの遷移の説明図である。

【図 29】

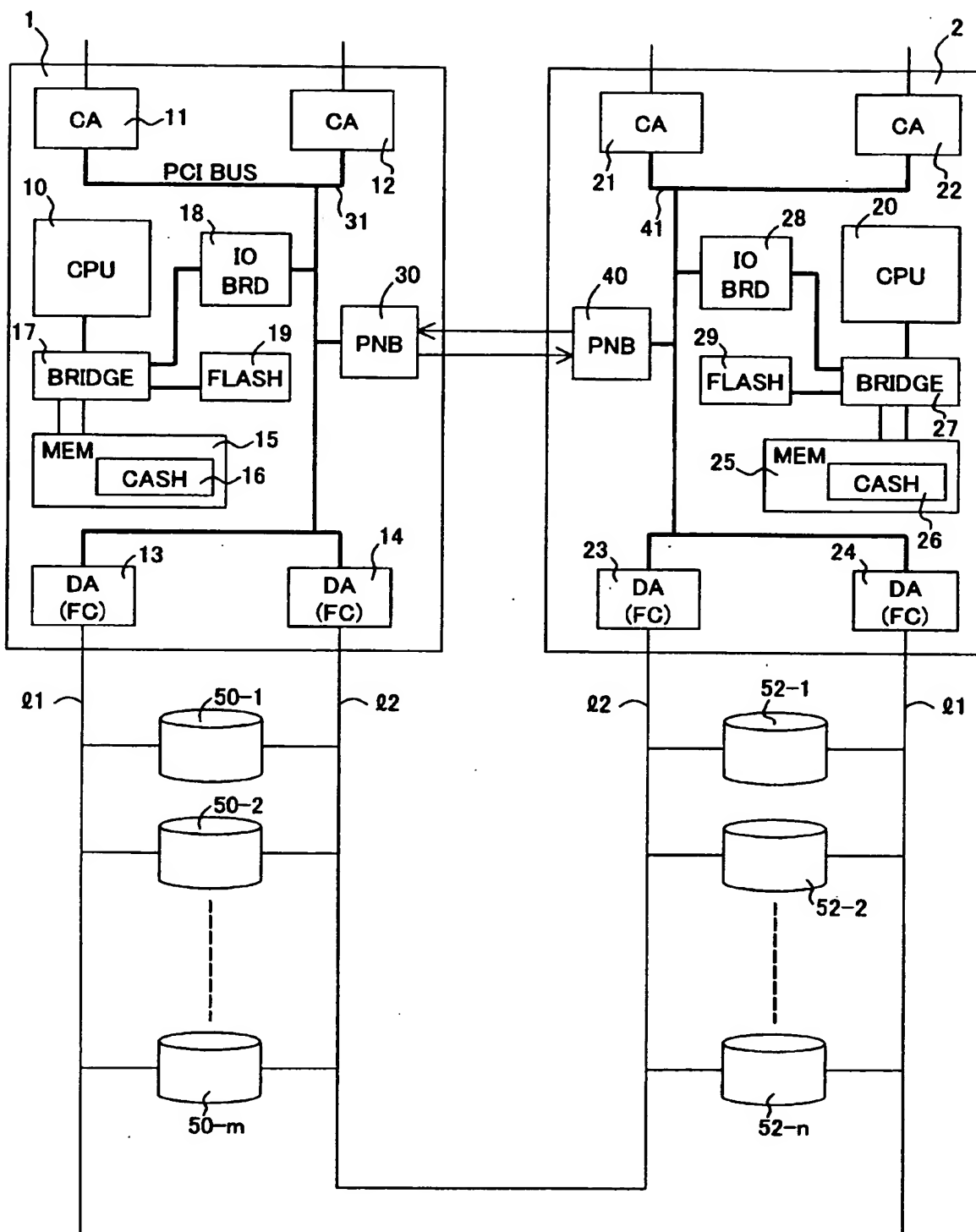
本発明の他の実施の形態のリード／ライトアドレス変換の説明図である。

【符号の説明】

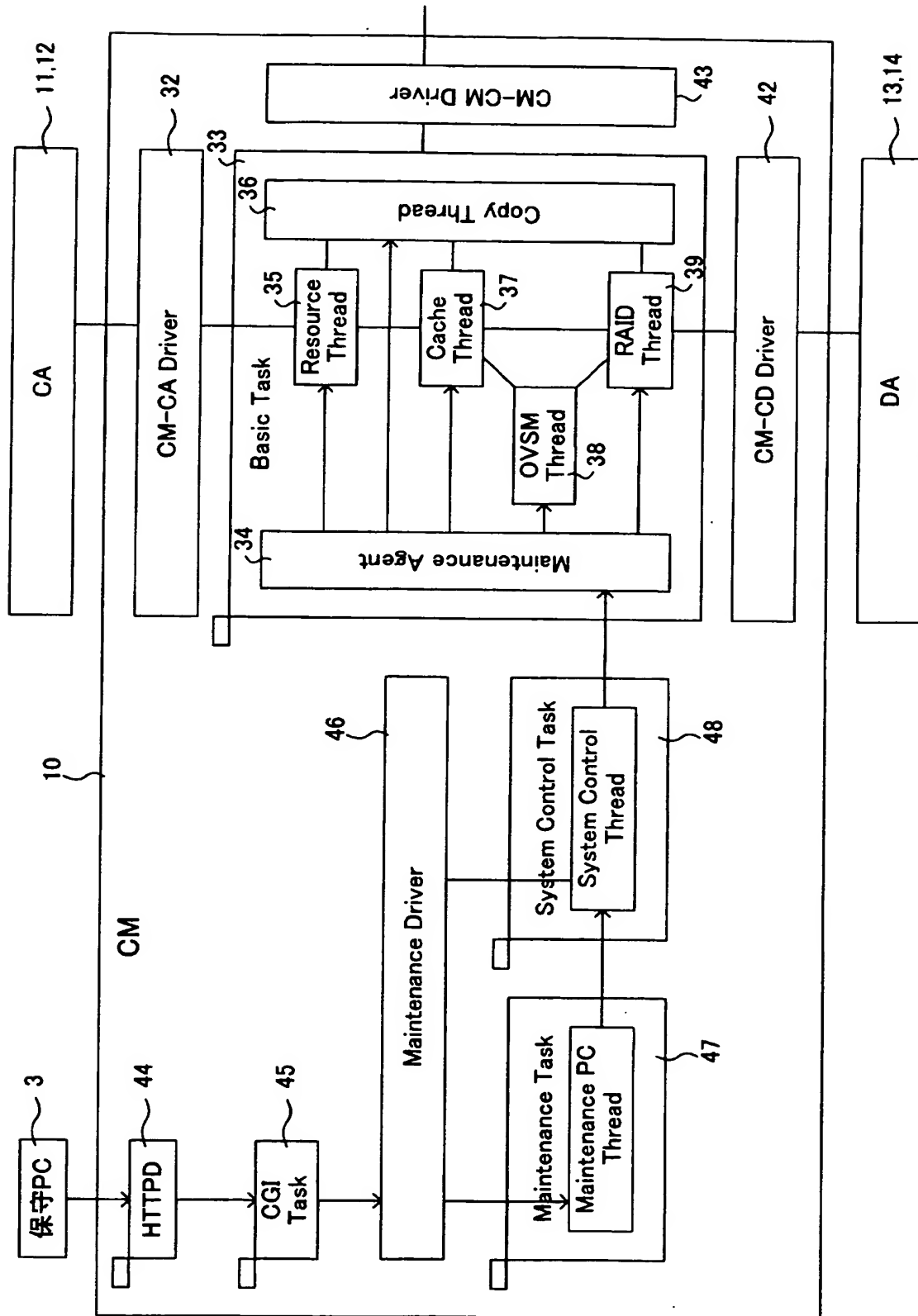
- 1、2 ストレージコントローラ
- 3 保守PC
 - 11、12、21、23 チャンネルアダプター
 - 13、14、23、24 デバイスアダプター
 - 10、20 CPU
 - 15、25 メモリ
 - 16、26 キャッシュメモリ
 - 30、40 PCI-ノードブリッジ回路
 - 31、41 PCIバス
 - 50、50-1～50-m、52-1～52-n 物理ディスク装置（ストレージ装置）
- 34 保守エージェント
- 35 リソーススレッド
- 37 キャッシュスレッド
- 38 OVSMスレッド
- 39 RAIDスレッド
- 48 システム制御スレッド
- 70 OLUテーブル
- 72 RLUテーブル
- 74 DLUテーブル
- 76 PLUテーブル
- 80 テンポラリRAID定義
- 82 RAID定義
- 90 キャッシュメモリ獲得処理
- 92 進捗状況管理処理
- 94 リード処理
- 96 ライト処理

【書類名】 図面

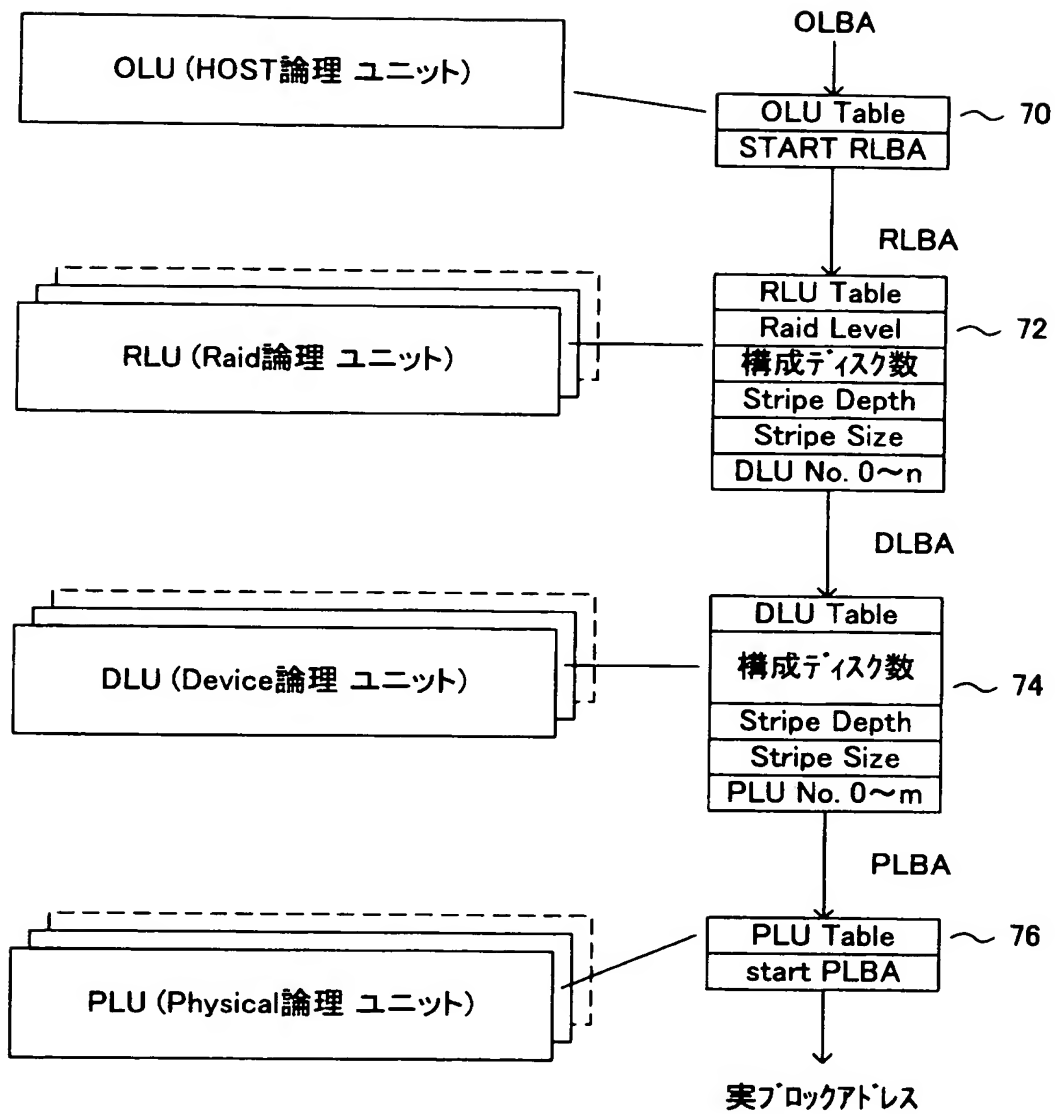
【図 1】



【図 2】

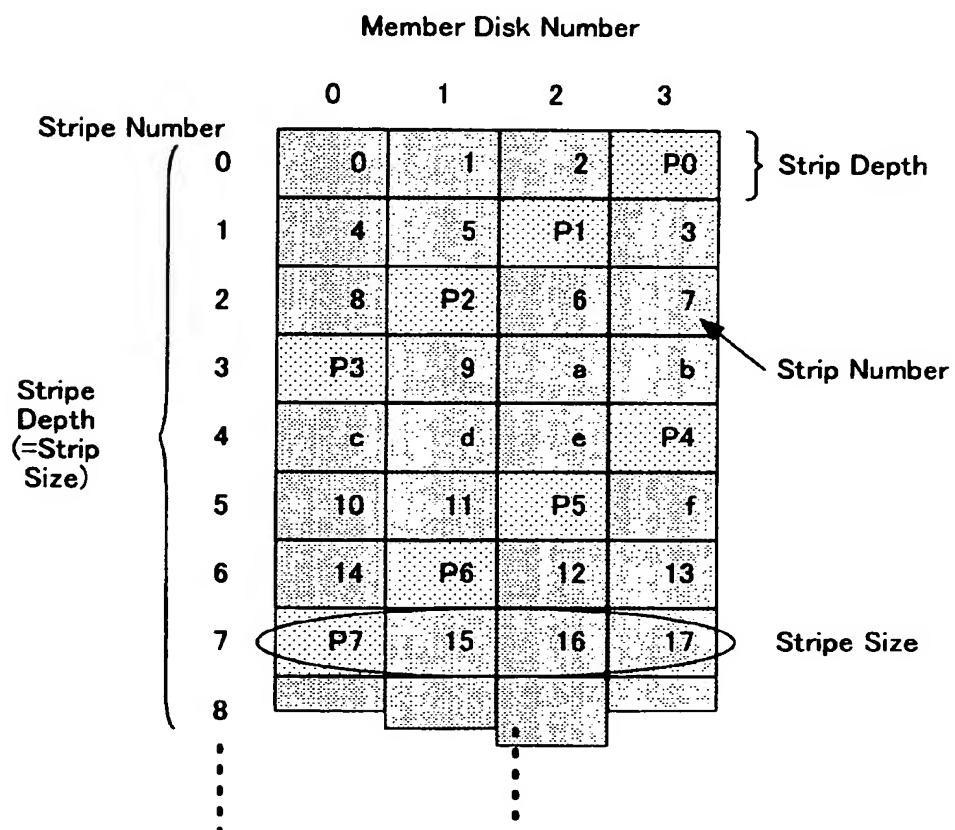
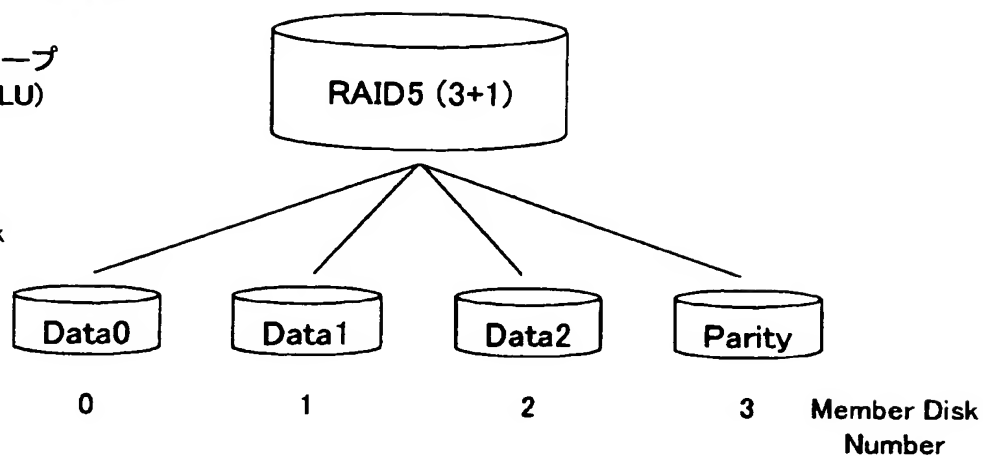


【図 3】

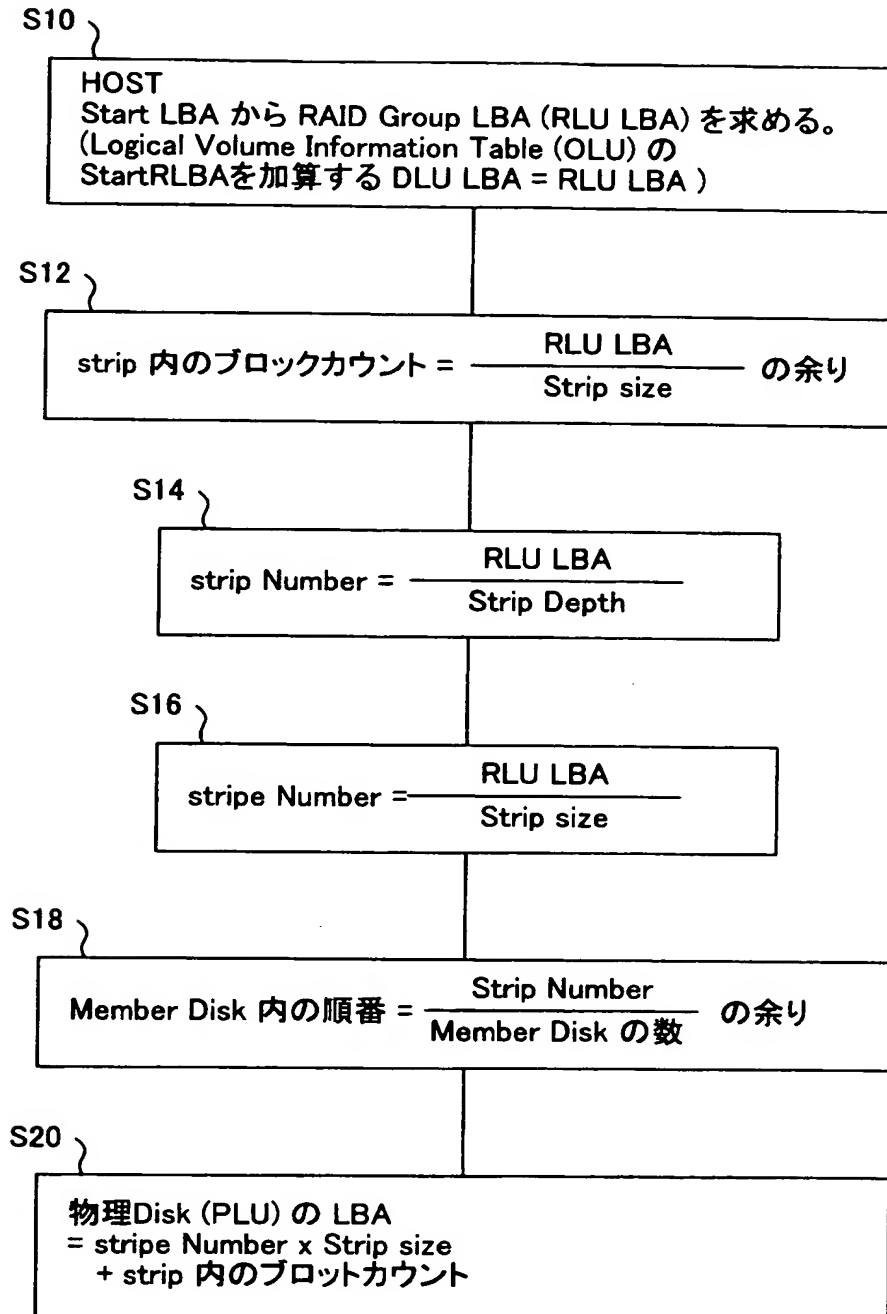


【図 4】

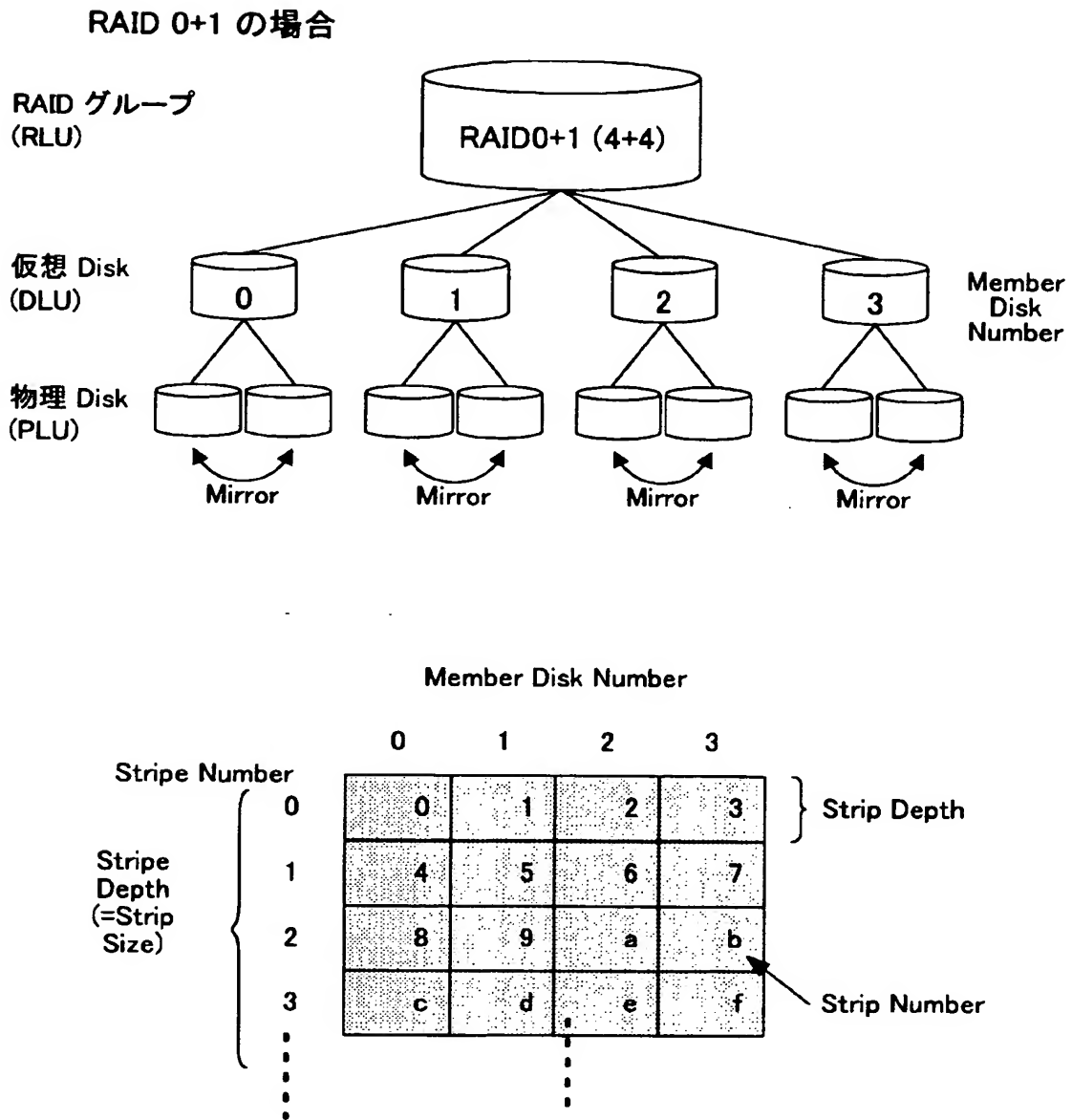
RAID 5 の場合

RAID グループ
(RLU = DLU)物理 Disk
(PLU)

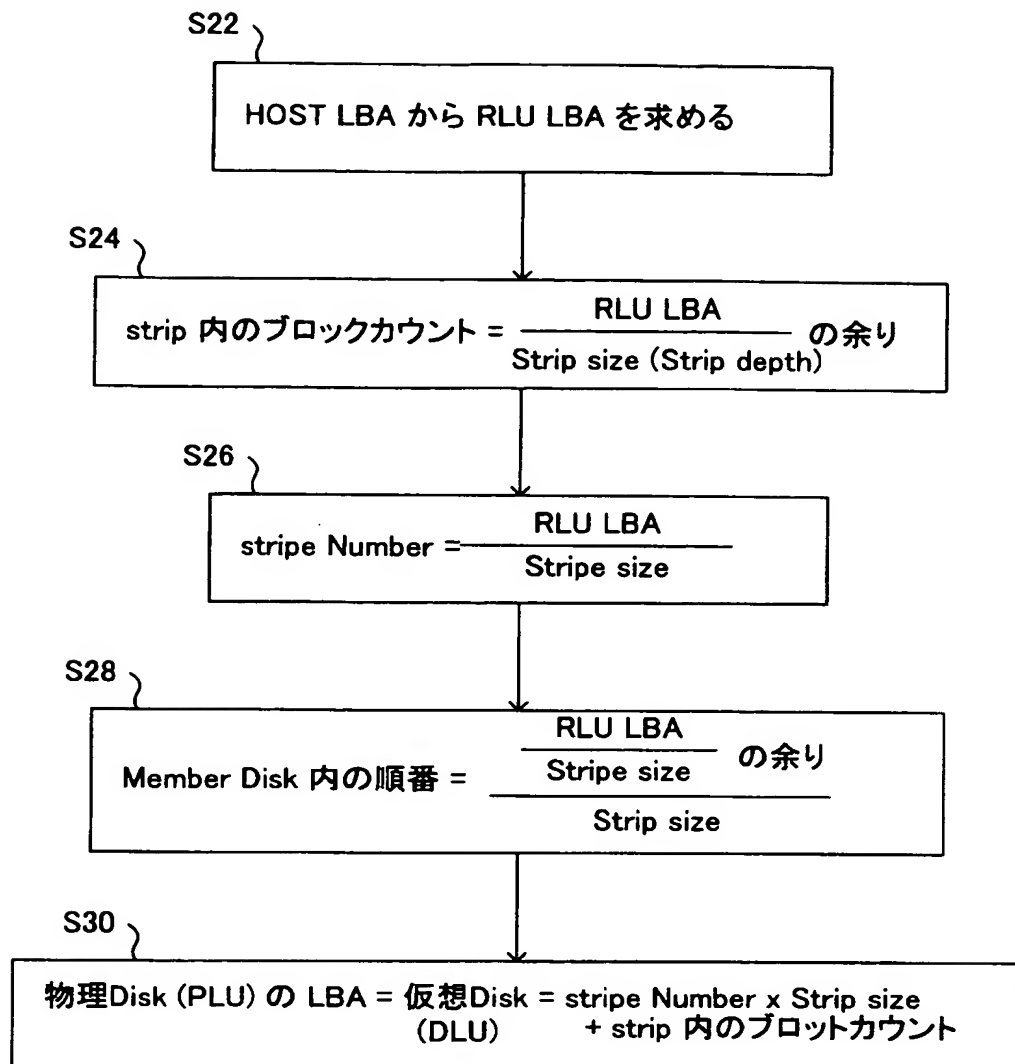
【図 5】



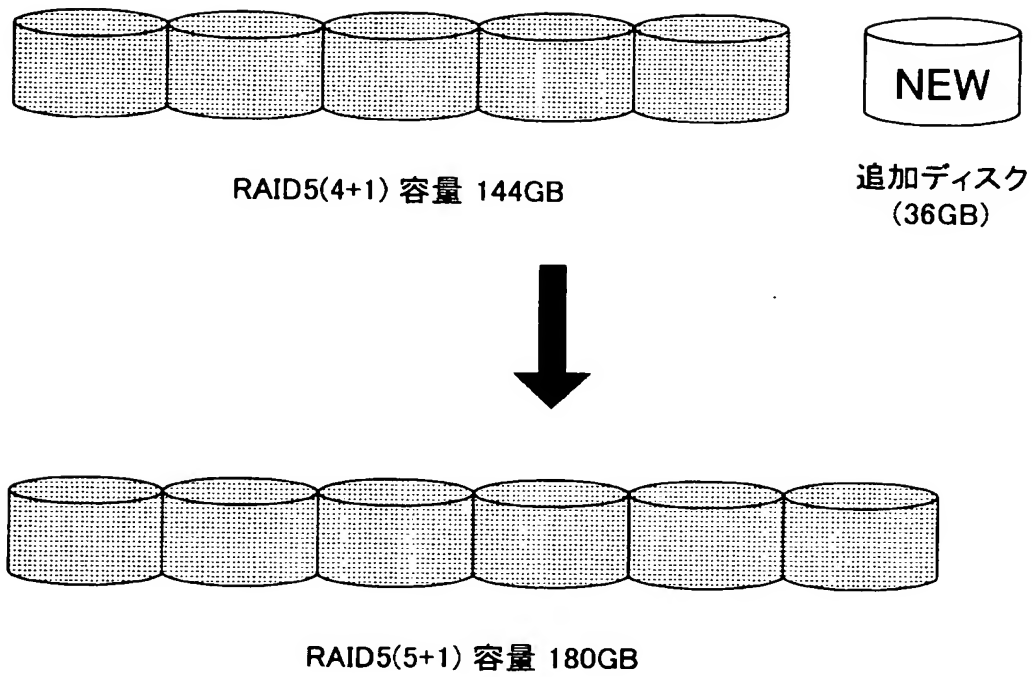
【図 6】



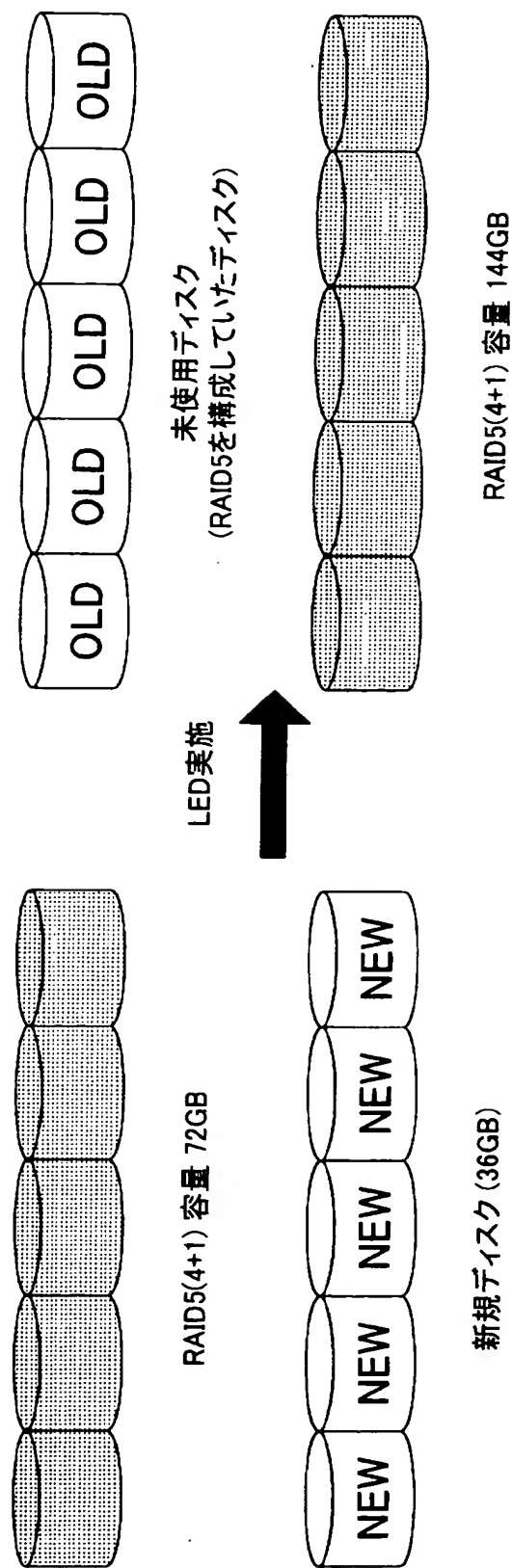
【図 7】



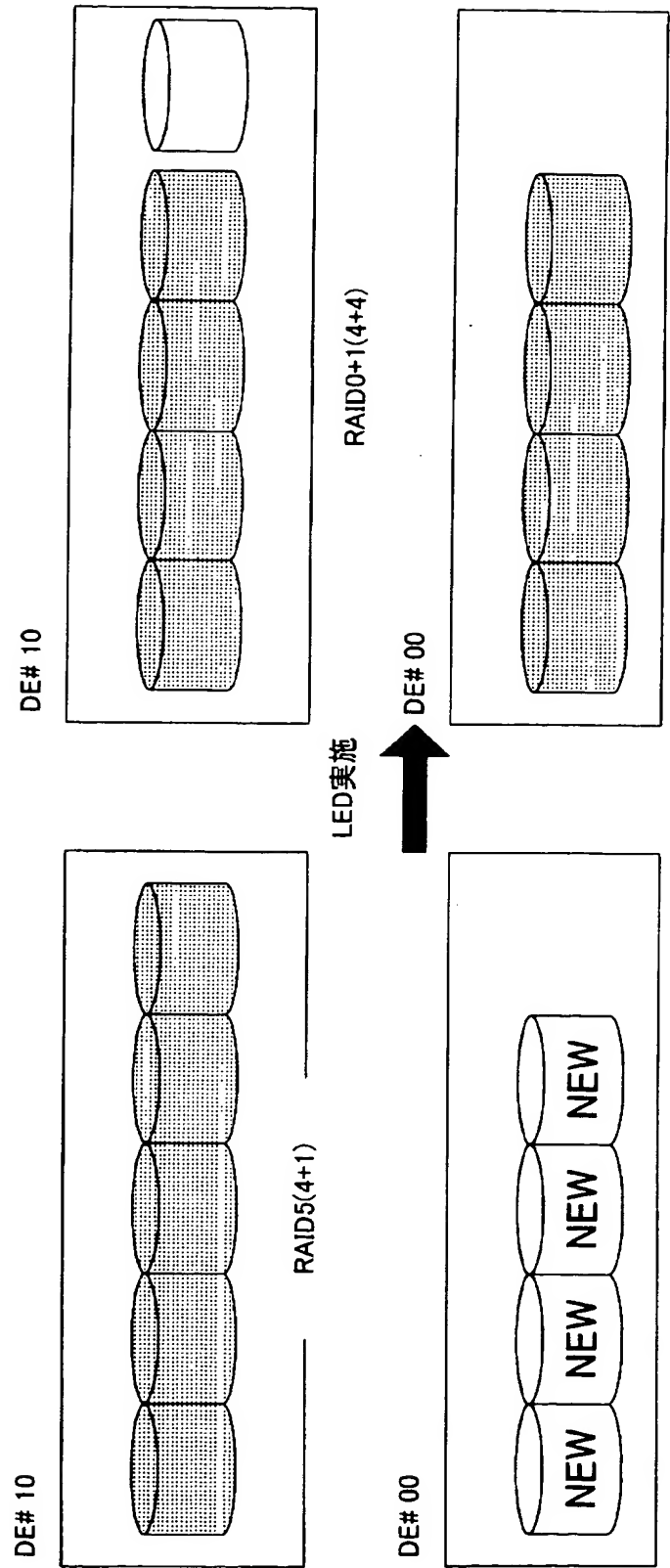
【図 8】



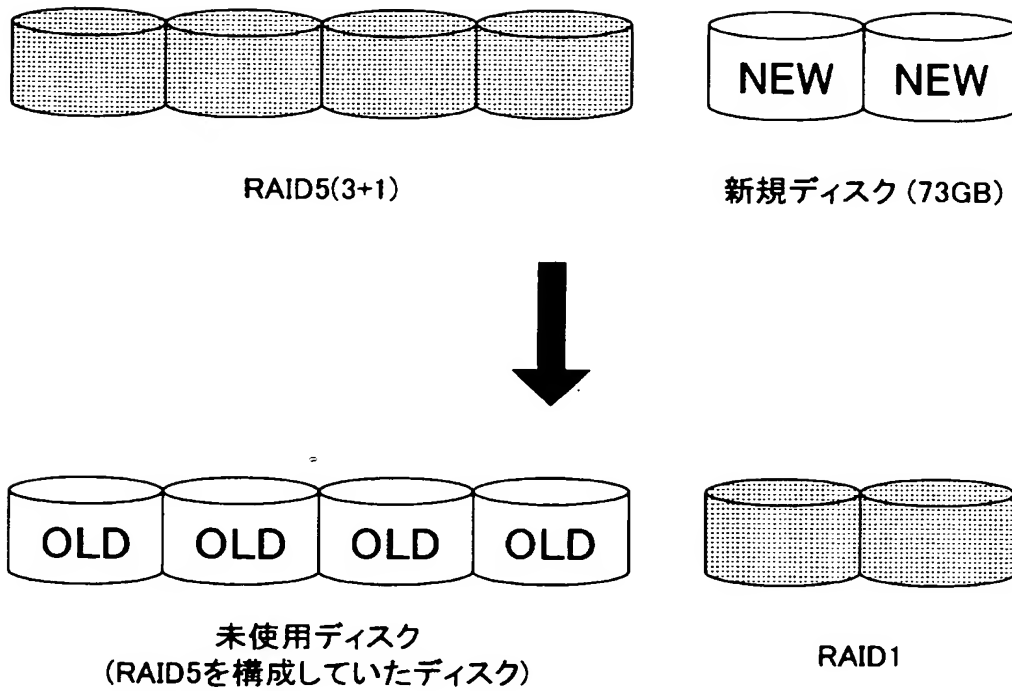
【図 9】



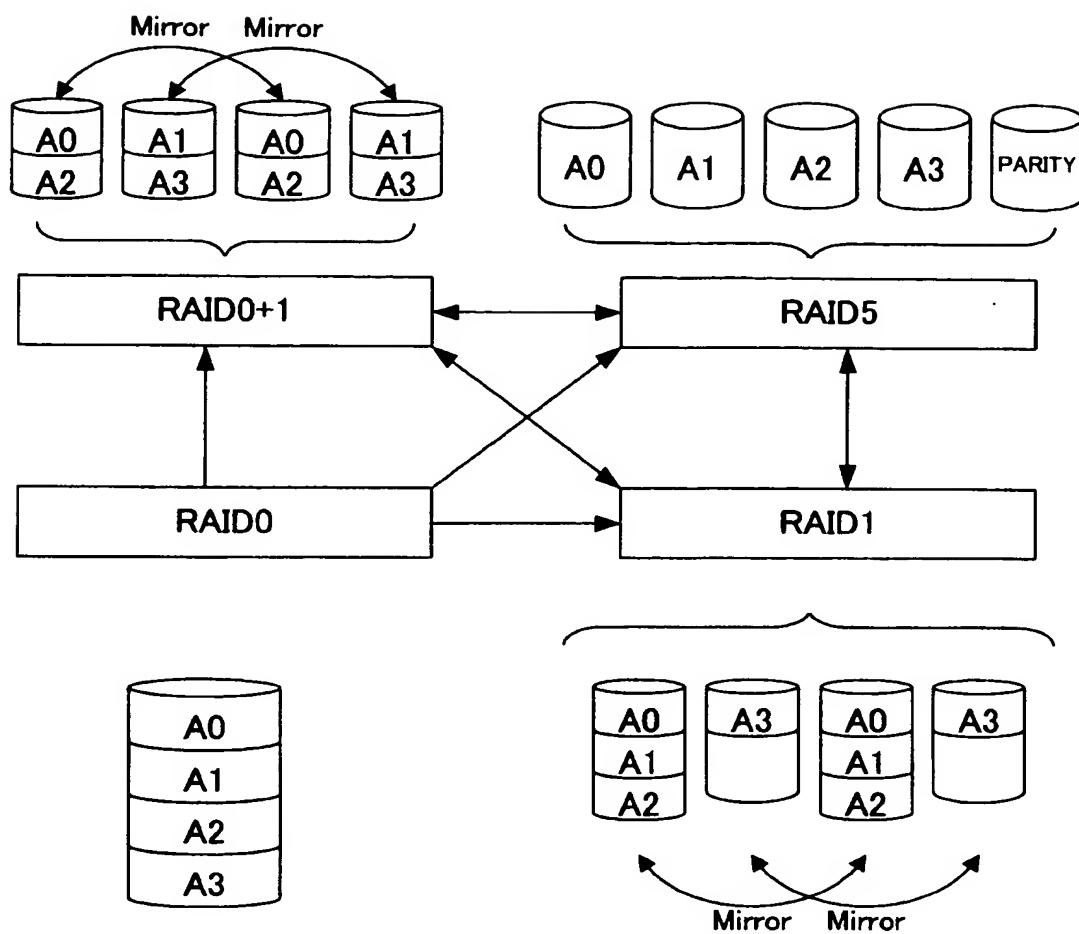
【図 1 0】



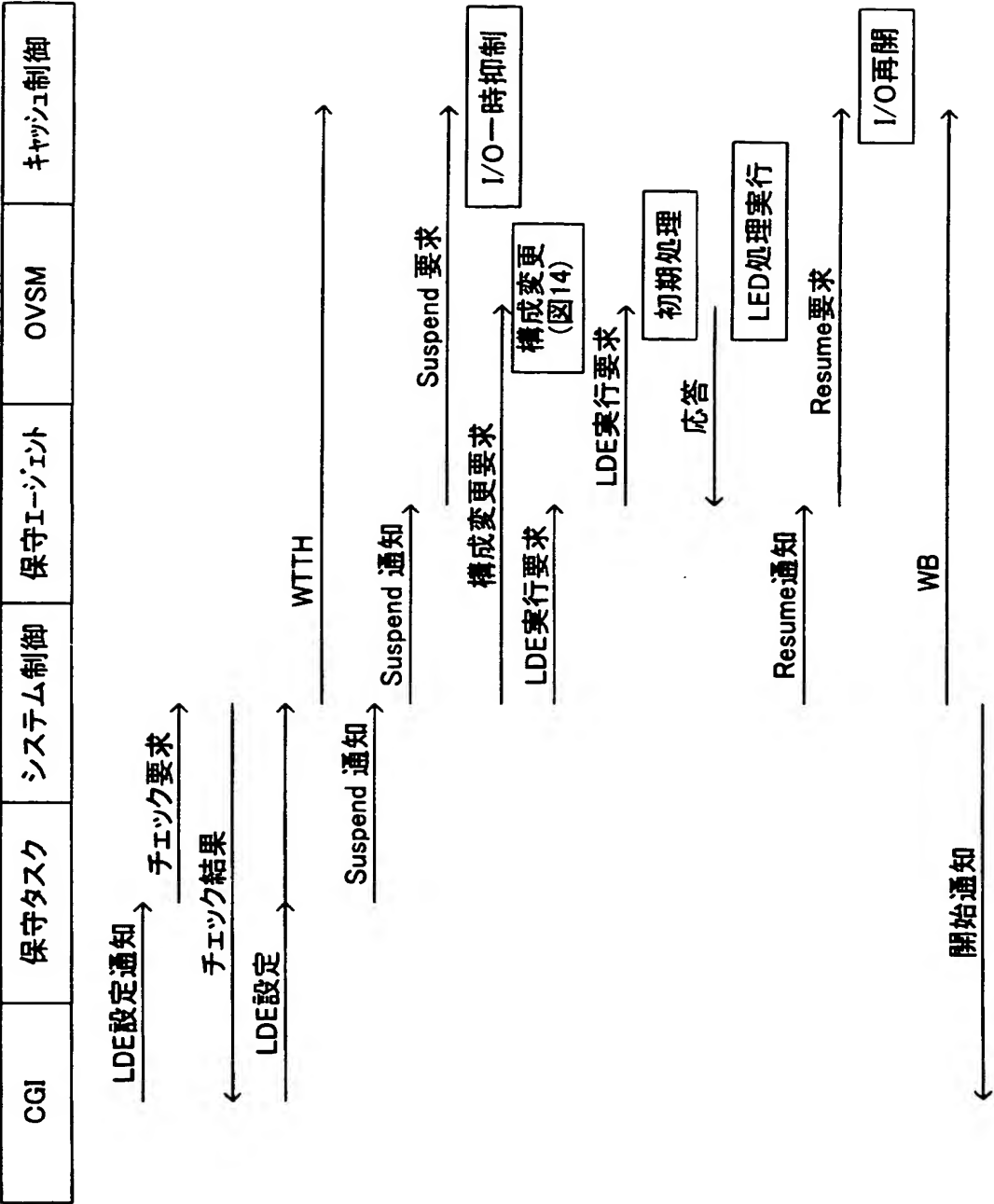
【図 11】



【図 12】



【図13】



【図 14】

	構成定義情報		LDE状況		
	LUN	項目	実施前	処理中	終了
RAID 定義 (新構成)	RLU			←	←
		Level	RAID5	←	←
		MemberDisks	5	6	←
		PLUNs	10-14	10-15	←
		StartLBA	0	←	←
		BlockCount	400	500	←
		CheckCode	id#0	←	←
		CVM mode	0	LDE Flag	0
	DLU	DLUN	0	←	←
	PLU				

82

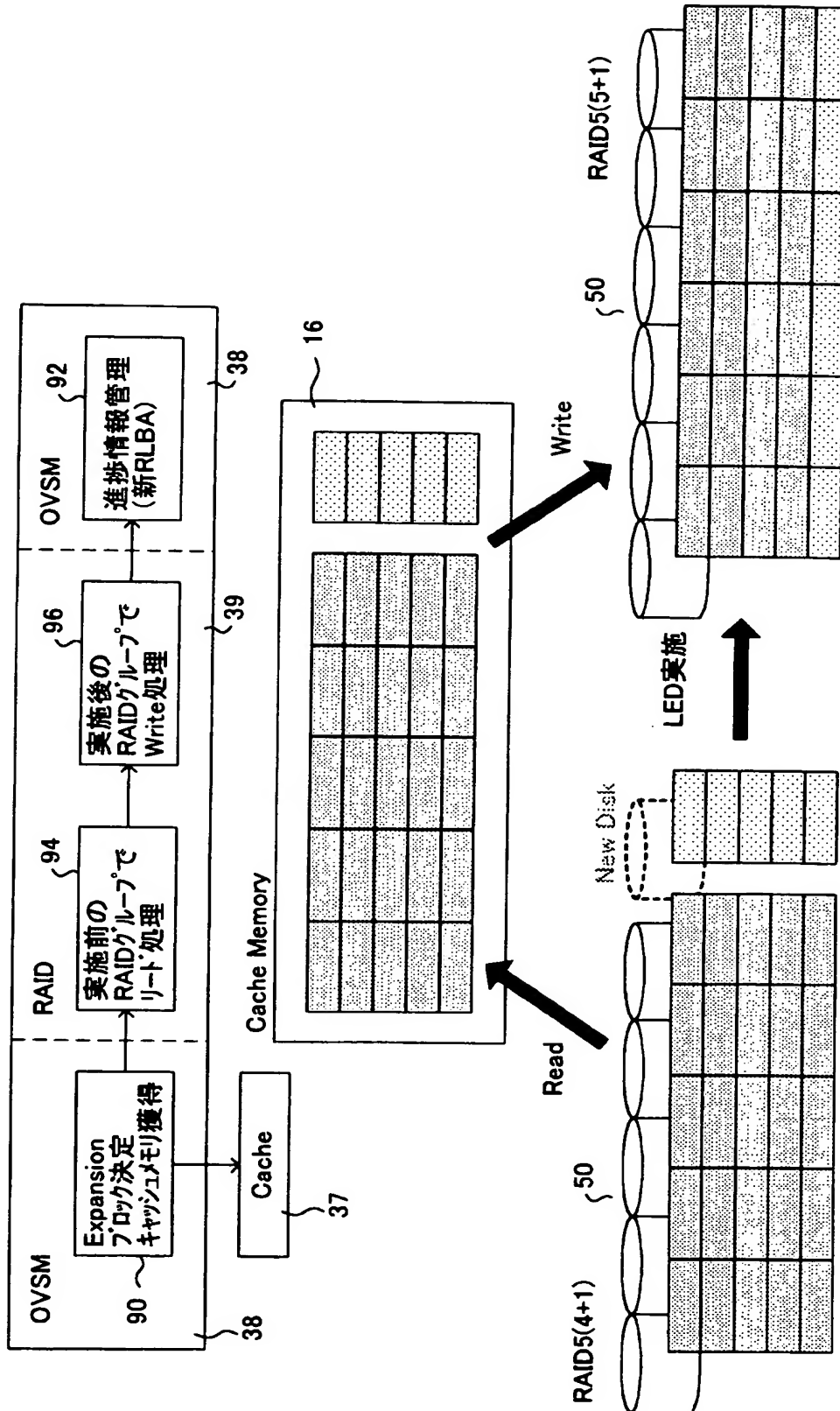
テンポラリ RAID 定義 (旧構成)	RLU			-	-
		Level	-	RAID5	-
		MemberDisks	-	5	-
		PLUNs	-	10-14	-
		StartLBA	-	0	-
		BlockCount	-	400	-
		CheckCode	-	id#0	-
		CVM mode	-	T Flag	-
	DLU	DLUN	-	255	-
	PLU	-	-	-	-

80

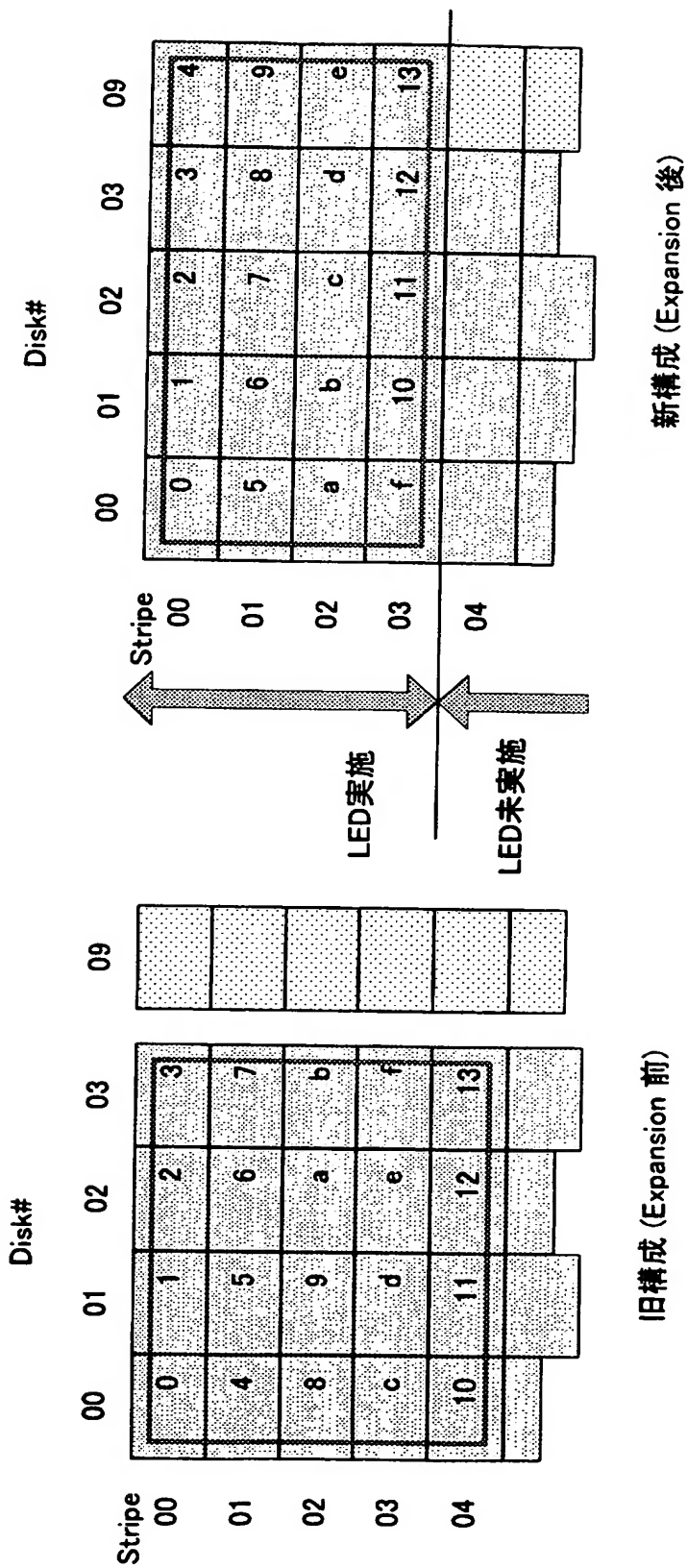
82

80

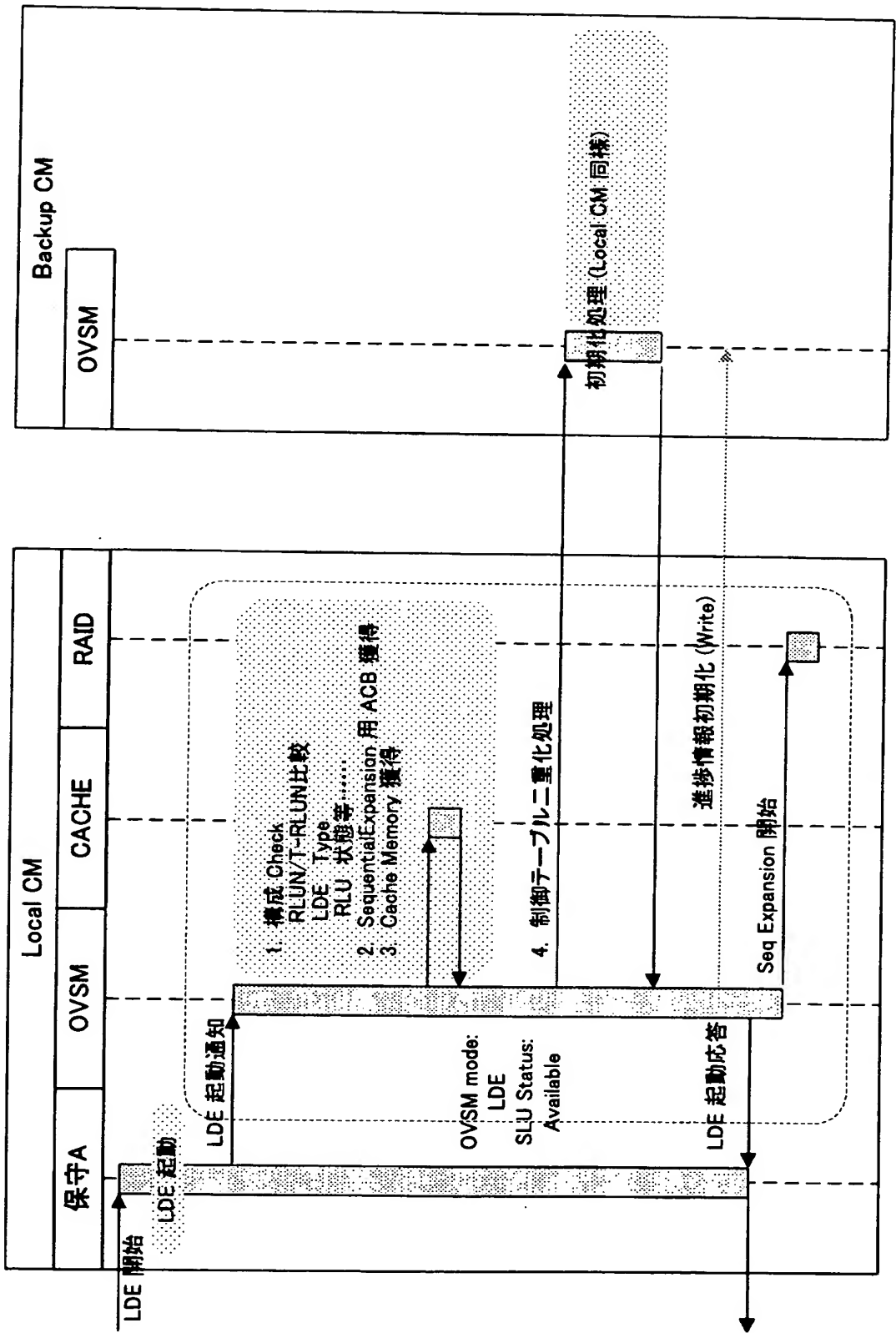
【図 15】



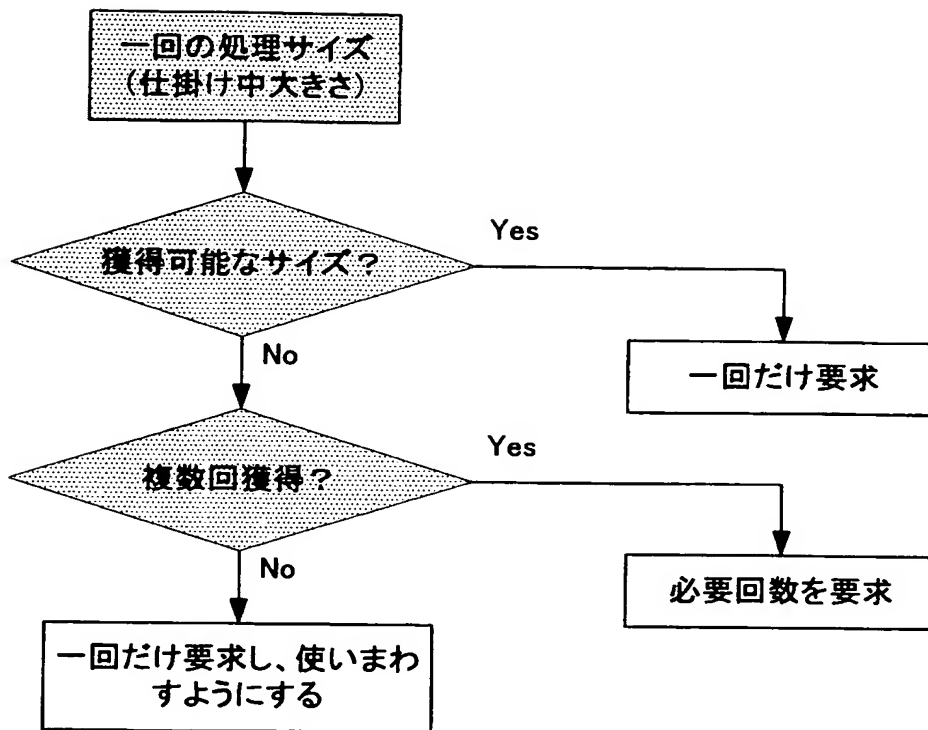
【図 16】



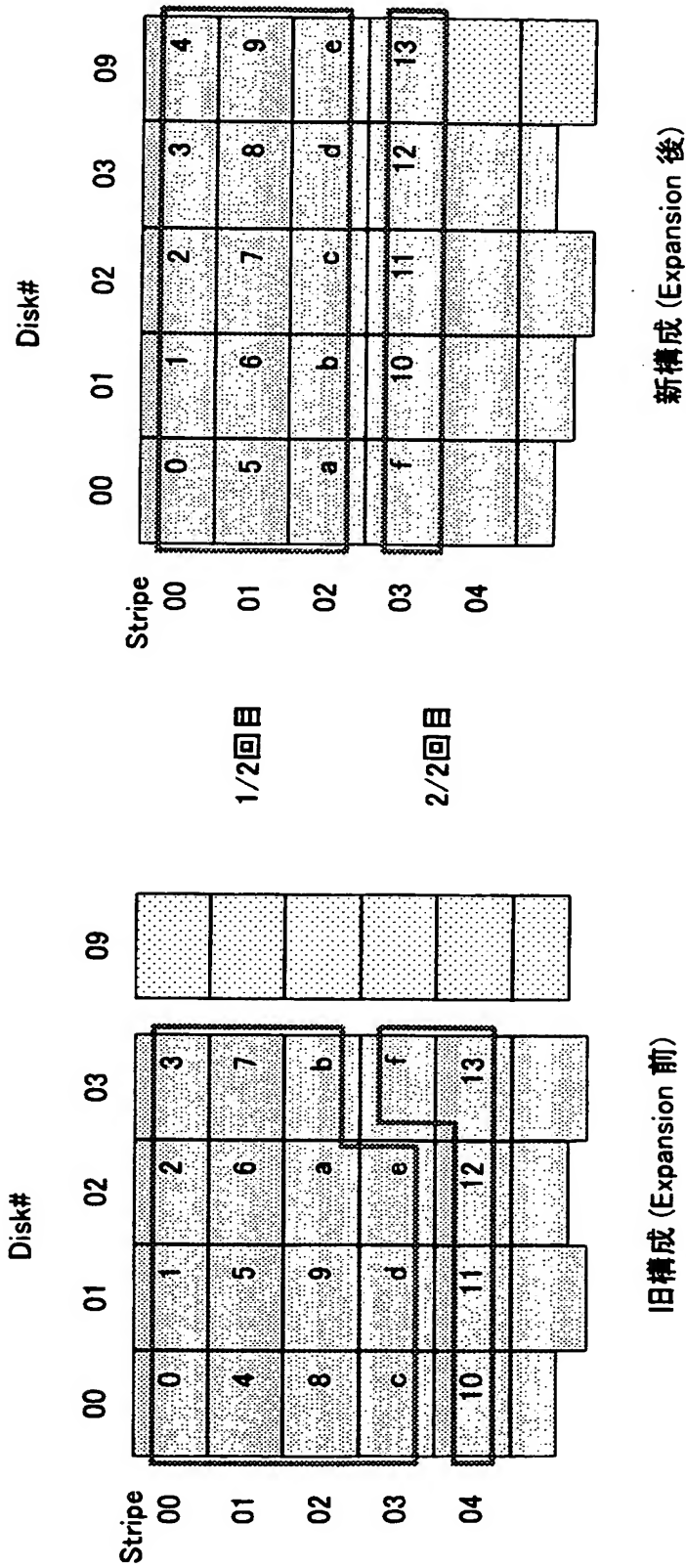
【図 17】



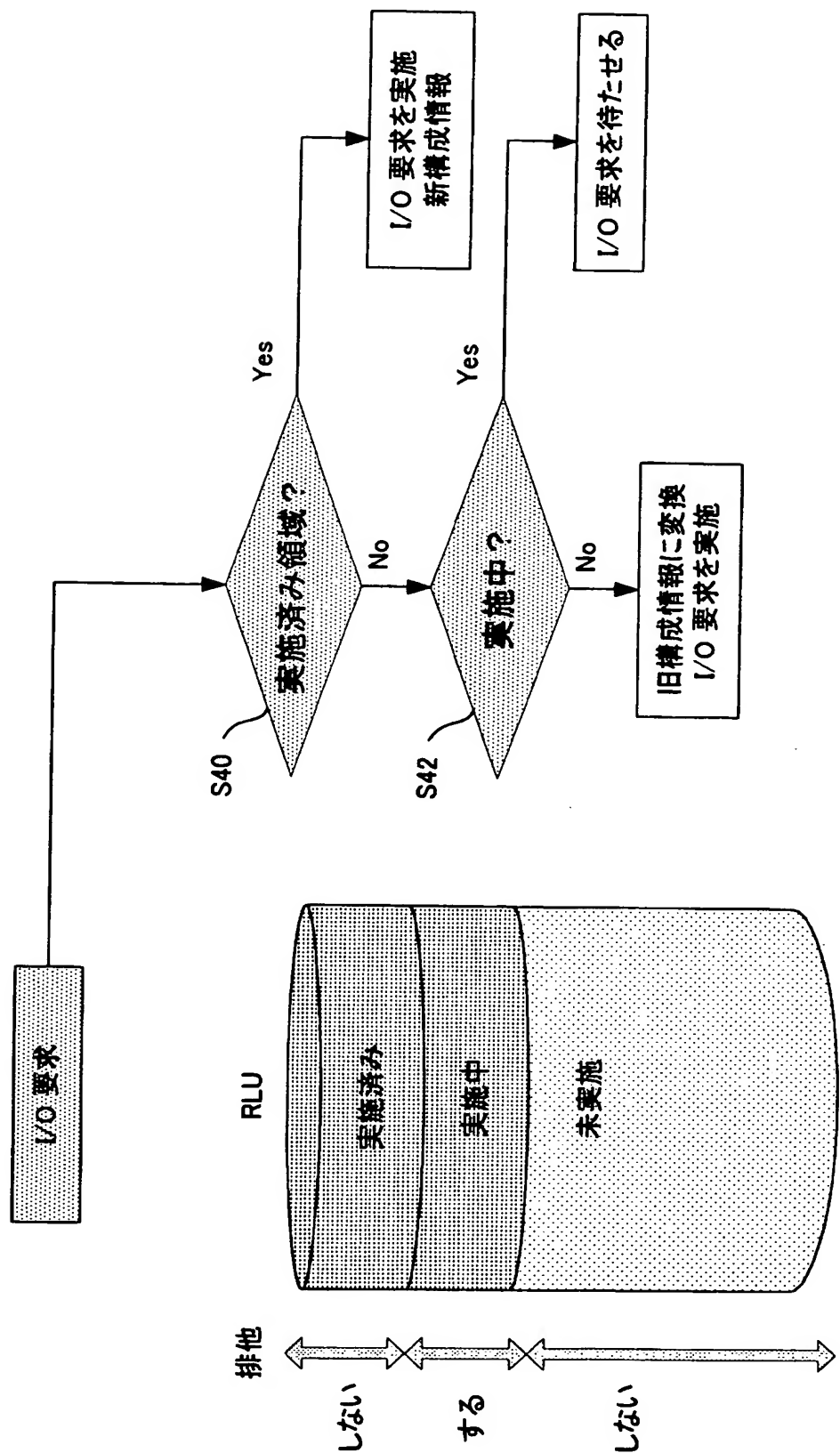
【図 18】



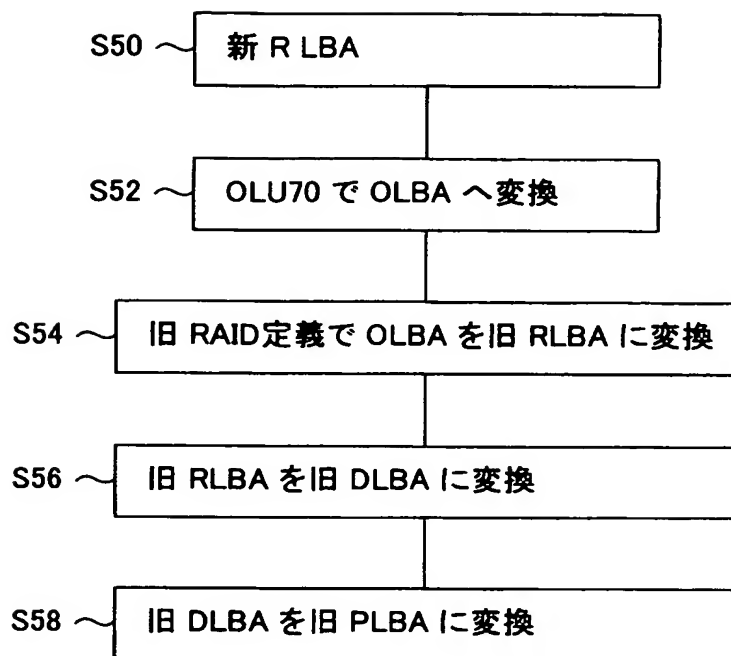
【図 19】



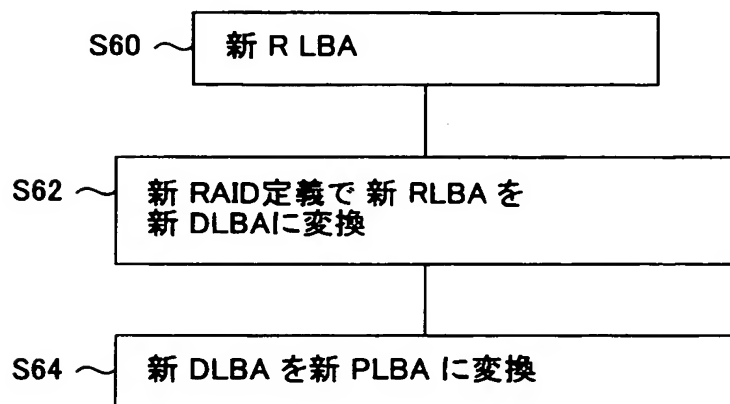
【図 20】



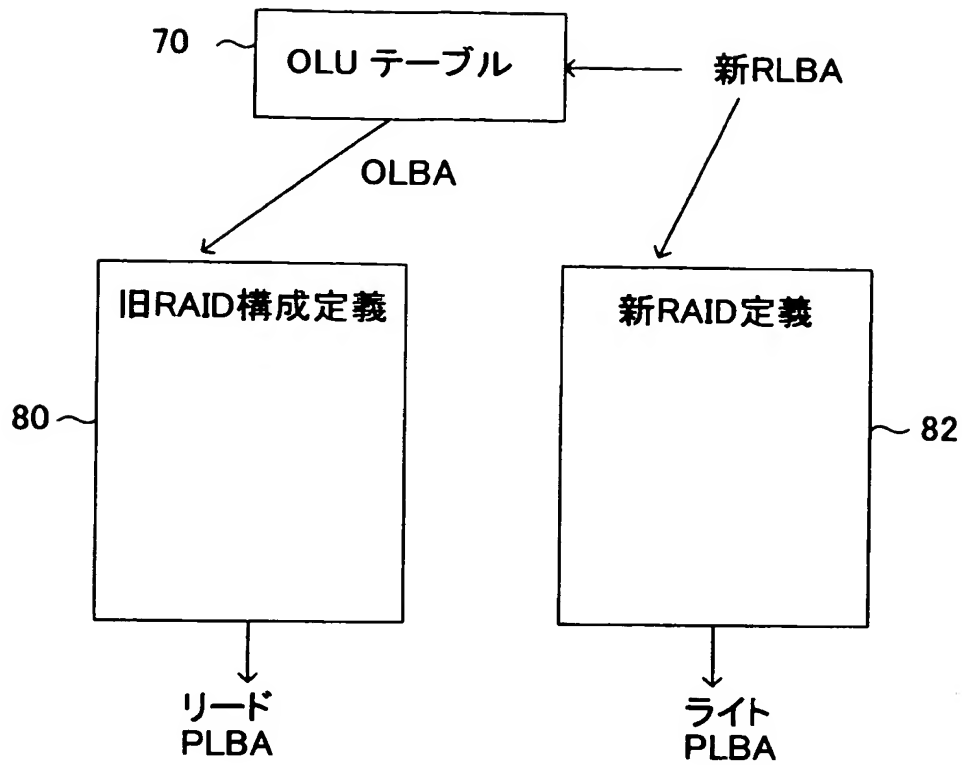
【図 2 1】



【図 2 2】



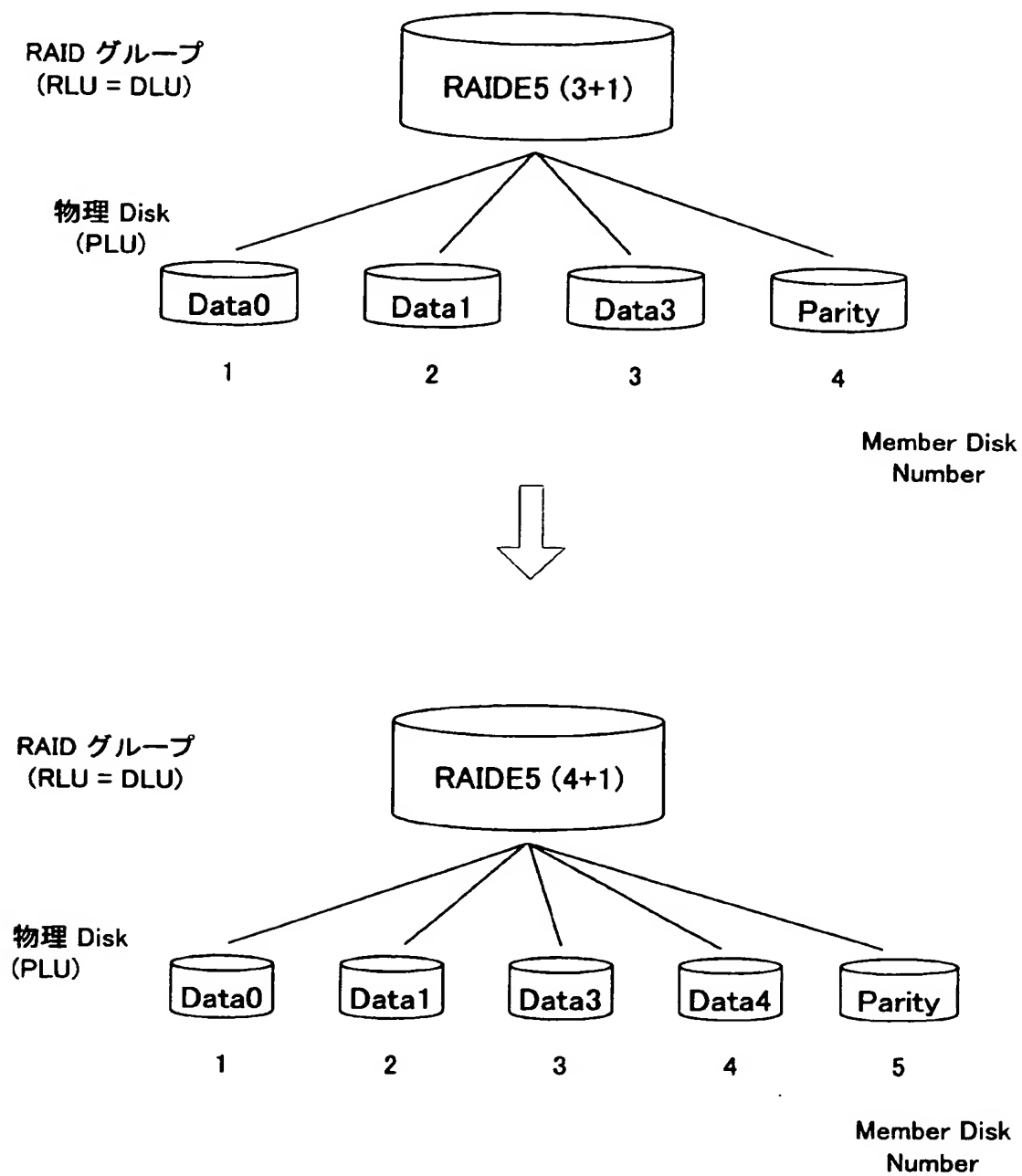
【図 23】



【図 24】

	構成定義情報		LDE状況		
	LUN	項目	実施前	処理中	終了
RAID 定義	RLU	RLUN	0	←	←
		Level	RAID5	←	←
		MemberDisks	4	5	←
		PLUNs	10-13	10-14	←
		StartLBA	0	←	←
		BlockCount	400	500	←
		CheckCode	id#0	←	←
		Status (VM mode)	0	LDE Flag	0
	DLU	DLUN	0	←	←
テンポラ RAID 定義	RLU	RLUN	-	255	-
		Level	-	RAID5	-
		MemberDisks	-	4	-
		PLUNs	-	10-13	-
		StartLBA	-	0	-
		BlockCount	-	400	-
		CheckCode	-	id#0	-
		Status (VM mode)	-	T Flag	-
	DLU	DLUN	-	255	-

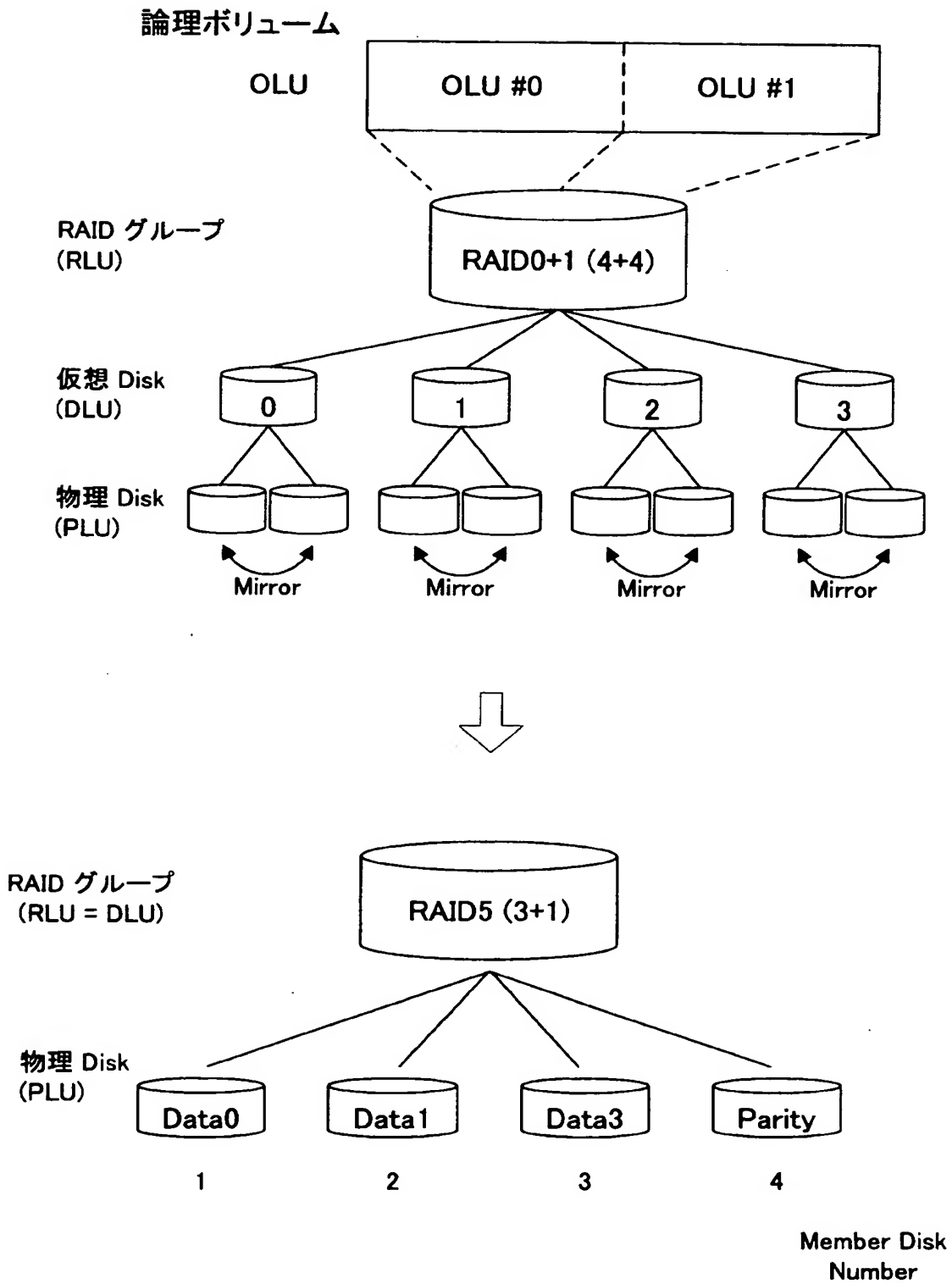
【図 25】



【図 26】

	構成定義情報		LDE状況		
	LUN	項目	実施前	処理中	終了
82 RAID 定義	RLU	RLUN	0	←	←
		Level	RAID0+1	RAID5	←
		MemberDisks		4	←
		PLUNs		0~3	←
		StartLBA	0	←	←
		BlockCount	400		←
		CheckCode	id#0	←	←
		CVM mode	0	LDE Flag	0
	DLU	DLUN	0	←	←
80 テンポラリ RAID 定義	RLU	RLUN	-	255	-
		Level	-	RAID0+1	-
		MemberDisks	-	4	-
		PLUNs	-	0~3	-
		StartLBA	-	0	-
		BlockCount	-	400	-
		CheckCode	-	id#0	-
		CVM mode	-	T Flag	-
	DLU	DLUN	-	255	-
	PLU	-	-	-	-

【図 27】



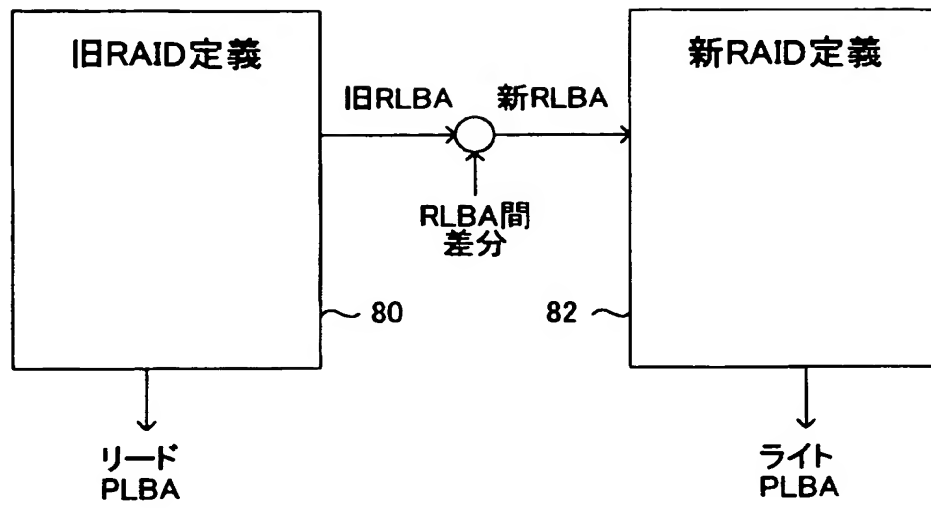
【図 28】

RLU		DLU 1	
RAID Level	0+1	Stripe Depth	128
Stripe Depth	128	Stripe Size	128
Stripe Size	128x4	Member Disk Count	2
Member Disk Count	4	DLU 2	
		DLU 3	
		DLU 4	



RLU		DLU	
RAID Level	5	Stripe Depth	128
Stripe Depth	128x3	Stripe Size	128x3
Stripe Size	128x3	Member Disk Count	4
Member Disk Count	1		

【図 29】



【書類名】 要約書

【要約】

【課題】 R A I D構成の冗長度を変更する R A I D装置において、多様な R A I Dレベル変換、容量増加を可能とする。

【解決手段】 少なくとも、R A I Dレベルと論理デバイス数を定義した新旧の R A I D構成定義情報（8 0、8 1）を使用し、制御部（1 0）が、それぞれにより R L Uマッピングして、ステージング、ライトバックして、R A I D構成を変更する。このため、多様な R A I Dレベルの変換、容量増加を実現できる。

【選択図】 図 1 5

特願 2 0 0 2 - 3 7 8 2 8 4

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 5 2 2 3]

1. 変更年月日

1 9 9 6 年 3 月 2 6 日

[変更理由]

住所変更

住 所

神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号

氏 名

富士通株式会社

特願 2 0 0 2 - 3 7 8 2 8 4

出 願 人 履 歴 情 報

識別番号

[0 0 0 1 3 6 1 3 6]

1. 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

石川県河北郡宇ノ気町字宇野気ヌ 9 8 番地の 2

氏 名

株式会社ピーエフユー

2. 変更年月日

2 0 0 3 年 4 月 7 日

[変更理由]

名称変更

住 所

石川県河北郡宇ノ気町字宇野気ヌ 9 8 番地の 2

氏 名

株式会社 P F U